AD A 232 089

FINAL REPORT:

Synergetic Multisensor Fusion

J.K. Aggarwal, Principal Investigator

(CVRC TR-91-4-68)

1 July 1987 - 30 September 1990

U. S. ARMY RESEARCH OFFICE

Contract No. DAAL03-87-K-0089

THE UNIVERSITY OF TEXAS AT AUSTIN

COMPUTER AND VISION RESEARCH CENTER

AUSTIN, TEXAS 78712

FINAL REPORT:

Synergetic Multisensor Fusion

J.K. Aggarwal, Principal Investigator

(CVRC TR-91-4-68)
1 July 1987 - 30 September 1990
U. S. ARMY RESEARCH OFFICE
Contract No. DAAL03-87-K-0089

Computer and Vision Research Center

The University of Texas at Austin

Austin, Texas 78712

512/471-3259

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED

MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
11/30/90

3. REPORT TYPE AND DATES COVERED
Final Report 7/1/87 - 9/30/90

4. TITLE AND SUBTITLE
Synergetic Multisensor Fusion

5. FUNDING NUMBERS
DAAL03-87-K-0089

6. AUTHOR(S)
J. K. Aggarwal

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Texas of Austin
Austin, Texas 78712-1084

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARO 25021.5-PH

11. SUPPLEMENTARY NOTES
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Synergetic multisensor fusion is the process of integrating information obtained from different sensing modalities in order to extract additional information that cannot be obtained by separately processing the signals from the different sensors. The development of a computer vision system using synergetic multisensor fusion is a complex task which encompasses sensor modeling, environment modeling, determining the analytic models used to interrelate the different sensing mechanisms, determining the models used to interrelate the sensed parameters of imaged objects (such as thermal emissivity, visual reflectance, and radar reflectance), and devising algorithms to exploit the derived models. We have developed powerful and robust algorithms for computer vision tasks based upon synergetic multisensor fusion. Our approach is suitable for applications such as object recognition, tracking, surveillance, and autonomous guidance.

14. SUBJECT TERMS
sensor fusion, multisensor, computer vision, range images, intensity images, ladar images, thermal images, video images, expert system, outdoor scenes, passive aerial navigation

15. NUMBER OF PAGES
52

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std Z39-18
298-102

SYNERGETIC MULTISENSOR FUSION

FINAL REPORT

J. K. AGGARWAL, PRINCIPLE INVESTIGATOR

SEPTEMBER 30, 1990

U. S. ARMY RESEARCH OFFICE

CONTRACT DAAL03-87-K-0089

COMPUTER AND VISION RESEARCH CENTER
THE UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TEXAS 78712-1084

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

SYNERGETIC MULTISENSOR FUSION

Summary

Synergetic multisensor fusion is the process of integrating information obtained from different sensing modalities in order to extract additional information that cannot be obtained by separately processing the signals from the different sensors. The development of a computer vision system using synergetic multisensor fusion is a complex task which encompasses sensor modeling, environment modeling, determining the analytic models used to interrelate the different sensing mechanisms, determining the models used to interrelate the sensed parameters of imaged objects (such as thermal emissivity, visual reflectance, and radar reflectance), and devising algorithms to exploit the derived models. We have developed powerful and robust algorithms for computer vision tasks based upon synergetic multisensor fusion. Our approach is suitable for applications such as object recognition, tracking, surveillance, and autonomous guidance.

DTIC COPY INSPECTED 1

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

Final Report:

SYNERGETIC MULTISENSOR FUSION

Table of Contents

Statement of the Problem Studied.

The automated interpretation of sequences of images for the purposes of detecting man-made objects, recognition of objects, and locating and tracking objects has many important applications in the peace-keeping activities of the Department of Defense. These activities include automated surveillance and monitoring, autonomous navigation for smart weapons, and industrial robotics, among others. The extraction of useful information from digitized imagery in a timely manner is crucial to these tasks. Due to the large amount of data to be processed, the presence of noise in the imagery, the absence of complete information, the ill-posed nature of the problems, and inadequate modelling of the scene and the sensors, such extraction is a very complex task. A broad program of research in machine vision is needed to establish useful and practical methods for machine perception of targets and guidance of payloads.

Past research in machine perception has focused mainly on use of a single sensing modality, such as a video camera or an infrared camera. A great deal of effort has been devoted to interpreting imagery sensed by each (single) modality. The many techniques based on this approach work only in highly constrained environments and require enormous amounts of computational resources. Such research efforts in multisensor fusion are of limited scope, and consisted mainly of developing empirical, ad hoc techqniques to accomplish narrowly defined tasks. These efforts used simple strategies such as merely combining the results obtained from separately processing sensor outputs to produce a larger set of features to classify. This type of sensor fusion does not produce features that are more discriminatory, but simply increases the dimensionality of the feature space, at the cost of a sharp increase in the computational demands--thus nullifying the advantages of sensor fusion.

Multisensor fusion allows for the integration of information in a synergetic manner--that is, it allows the extraction of new information that cannot be obtained by separately processing signals from the various sensors. This characteristic of multisensor fusion is illustrated by the case of stereoscopy, where the determination of depth information is possible only when features from both left and right images are examined concommitantly. Such synergetic processing can be extended to the integration of diverse sensing modalities, e.g., the integration of information from thermal and visual images in order to obtain an estimation of surface heat fluxes at the surface of an object, which in turn yields features that enable object recognition. Synergetic multisensor fusion offers several other advantages, including an increase in the number of feature obtained, which leads to increased ability to discriminate objects; an increased robustness, due to the redundancy of information obtained, which provides fault-tolerance and the ability to adapt to changing conditions; and an increase in computational efficiency, since the additional information provided by synergetic multisensor fusion and the resulting increase in feature discrimination allow the use of simpler classification algorithms and provide increased accuracy in classification.

The primary objective of our research under this contract was to develop powerful and robust algorithms for computer vision tasks based on synergetic multisensor fusion. Our approach was intended to facilitate the integration of several sensing modalities, such as infrared, visual (including color and stereoscopy), radar, and other available active and passive sensing modes. In the course of developing such algorithms, it was necessary to identify and address the various issues involved in integrating sensing modalities, including the fundamental issues of sensor modeling, environment modeling, determining analytical models to interrelate the sensing mechanisms, determining models to interrelate the sensed parameters of the imaged object, such as

thermal emissivity, visual reflectance, and radar reflectance, and finally, determining the algorithms that make optimal use of the derived models. The work included establishing analytical models for sensors and environment, using these models to specify sensitive and specfic features, and devise algorithms based on these features to detect, classify, and track objects in the sensed scene.

Summary of Important Results.

Presentation of our research results has been organized as follows.

1. Multisensor Fusion of Thermal and Visual Images
2. Segmentation and Understanding of Ladar Images
3. Interpretation of Range Imagery
4. Passive Aerial Navigation by Image Sequence Analysis
5. Identifying Man-made Objects in Outdoor Scenes / Fusion of Color and Geometry Information
6. Positional Estimation Techniques for An Outdoor Mobile Robot

1. INTERPRETATION OF THERMAL AND VISUAL SENSOR INFORMATION.

Past research in computer vision has shown that the interpretation of a single image of a scene is a highly underconstrained task. Fusion from multiple cues from the same image, and fusion from multiple views using the same modality have been marginally successful. Recently, the fusion of information from different modalities of sensing has been studied to further constrain the interpretation. A number of approaches have been developed for image segmentation and the analysis and understanding of the segmented images using multisensor fusion. In this section, we describe a system that uses registered thermal and visual images for surface heat flux analysis, and an image synthesis system that generates visual images as well as thermal images based on internal heat flow in objects. In the following section, we detail a system based upon fusion of ladar (laser radar) and thermal images.

Our approach is based on the phenomenological modeling of infrared and visual signal generation and detection, and the relationship between these signals and the intrinsic thermal properties of the imaged objects. The approach combines information from thermal and visual imagery to classify objects based on the estimated lumped thermal capacitance of the imaged objects.

To briefly describe our approach developed for combining thermal and visual sensors for outdoor scene perception, we develop a computational model that allows us to derive a map of heat sinks and sources in the imaged scene based on estimates of surface heat fluxes. A feature which quantifies the surface's ability to sink/source heat radiation is derived. Aggregate region features are used in a decision tree based classification scheme to label image regions as vehicle, building, vegetation or road. Real data are used to illustrate the usefulness of the approach.

We assume that the thermal image is segmented into closed regions by a suitable segmentation algorithm and that the thermal and visual images are registered. The thermal image is processed to yield estimates of object surface temperature [1]. The visual image, which is spatially registered with the thermal image, yields information regarding the relative surface orientation of the imaged object [1]-[3]. This information is made available at each pixel of the images. Other information such as ambient temperature, wind speed, and the date and time of image acquisition is used in estimating the surface heat fluxes at each pixel of the image.

Consider an elemental area on the surface of the imaged object. Assuming one-dimensional heat flow, the heat exchange at the surface of the object is represented by Figure 1, where W_i is the incident solar radiation, q_i is the angle between the direction of irradiation and the surface normal, the surface temperature is T_s, and W_{abs} is that portion of the irradiation that is absorbed by the surface. W_{cv} denotes the heat convected from the surface to the air which has temperature T_{amb} and wind speed V, W_{rad} is the heat lost by the surface to the environment via radiation and W_{cd} denotes the heat conducted from the surface into the interior of the object.

At any given instant, applying the principle of conservation of energy at the surface, the heat fluxes flowing into the surface of the object must equal those flowing out from the surface. We therefore have

$$W_{abs} = W_{cv} + W_{cd} + W_{rad} \quad (1)$$

W_{abs} is computed at each pixel using surface reflectivity and relative surface orientation information which is estimated from the visual image, along with knowledge of the incident solar radiation. W_{rad} is computed from Stefan-Boltzman's law knowing sky temperature and surface temperature. We use the empirical convection correlations developed for external flow over flat plates for computing W_{cv}. Having estimated W_{abs}, W_{cv} and W_{rad}, W_{cd} is estimated using equation (1). The estimation of surface heat fluxes is described in detail in references [1]-[3].

The surface heat balance described by equation (1) and Figure 1 depends on several time varying parameters. In such a dynamic situation the rate of heat loss / gain at the surface must equal the rate of change of internal energy of the object. Hence, we have:

$$W_{cd} = DVc\ dT_s/dt$$

where D is the density of the object, V is the volume of the object, c is the specific heat of the object, and t denotes time. Let h denote the convection heat transfer coefficient, ε_o the surface emissivity, and σ the Stefan-Boltzman constant. Considering a unit surface area, the equivalent thermal circuit for the surface is shown in Figure 2, where the resistances are given by:

$$R_{cv} = 1/h\ , \quad R_{rad} = 1/\varepsilon_o\sigma(T_s^2 + T_{amb}^2)(T_s + T_{amb})$$

Note the dependence of the latter on the driving potential, i.e., the temperature difference. The lumped thermal capacitance of the object is given by $C_t = DVc$. A relatively high value for C_t implies that the object is able to sink or source relatively large amounts of heat. Note that the conduction heat flux at the surface of the object is the component that affects the internal energy of the object, and is dependent upon both the rate of change of temperature as well as the thermal capacitance. In our experiments, we have found the rate of change of surface temperature to be very small except during the short period of time when the surface of the object enters into or exits from a shadow [1]. Hence, in general, the predominant factor in determining the conduction heat flux is the thermal capacitance of the object.

An estimate of W_{cd} provides us with a relative estimate of the thermal capacitance of the object, albeit a very approximate one. Table 1 lists values of C_t of typical objects imaged in outdoor scenes. The values have been normalized for unit volume of the object. The value shown for automobiles has been computed using the volume of an entire automobile, its weight, and the specific heat value for mild steel.

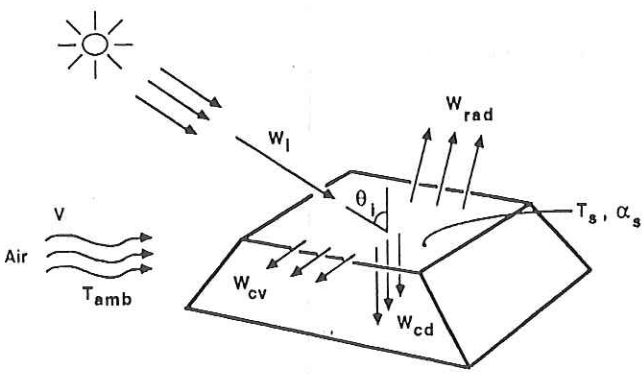

Figure 1. Surface heat fluxes

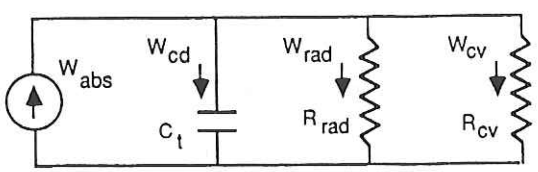

Figure 2. Equivalent thermal circuit of the imaged surface.

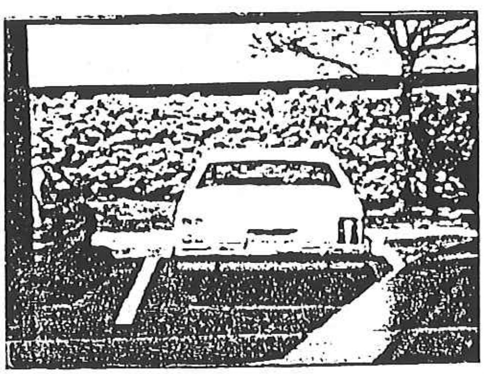

Figure 3. Visual Image of Scene

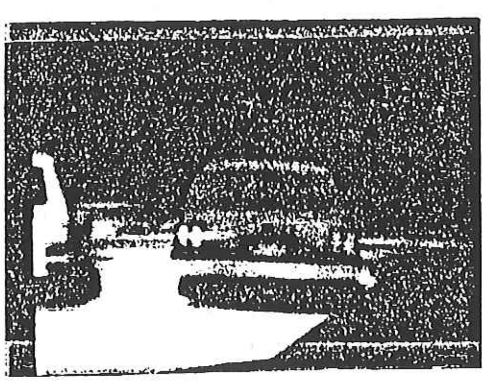

Figure 4. Thermal Image of Scene

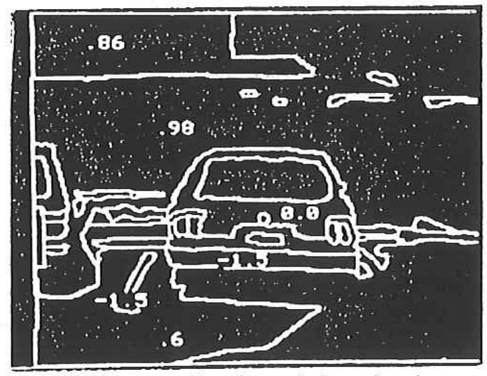

Figure 5. Mode of heat flux ratio for each region

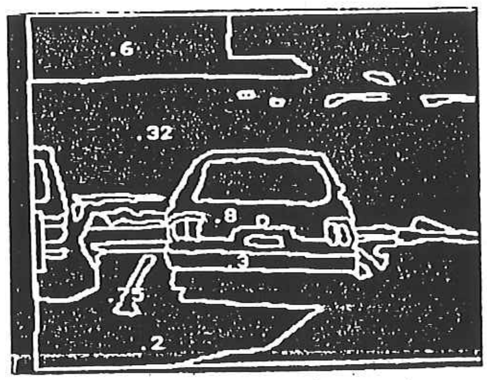

Figure 6. Surface reflectivity for each region

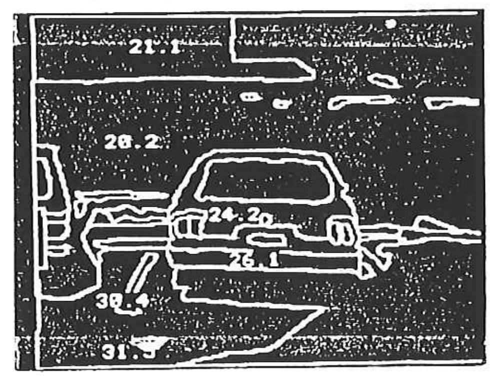

Figure 7. Average region temperature

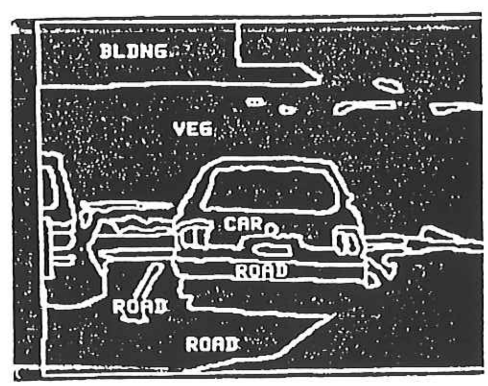

Figure 8. Region labelling by classifier

Object (x10^{-6} Joules/Kelvin)	***Thermal Capacitance***
Asphalt Pavement	1.95
Concrete Wall	2.03
Brick Wall	1.51
Wood(Oak) Wall	1.91
Granite	2.25
Automobile	0.18

Table 1. Normalized values of lumped thermal capacitance.

Note that the thermal capacitance for walls and pavements is significantly greater than that for automobiles; hence, W_{cd} may be expected to be higher for the former regions. Plants absorb a significant percentage of the incident solar radiation for photosynthesis and transpiration. Only a small amount of the absorbed radiation is convected into the air. Therefore, if equation (1) is used, the estimate of the W_{cd} will be almost as large (typically 95%) as that of the absorbed heat flux. Thus W_{cd} is useful in estimating the object's ability to sink/source heat radiation, a feature shown to be useful in discriminating between classes of objects. Note that W_{cd} is proportional to the magnitude of solar irradiation incident on that surface element. In order to minimize the feature's dependence on differences in absorbed heat flux, a normalized feature was defined to be the ratio $R = W_{cd}/W_{abs}$.

Although the heat flux ratio W_{cd}/W_{abs} captures a great deal of information about the imaged object, it cannot unambiguously identify the imaged object. Other sources of information must be used. Our classification scheme uses information such as the surface reflectivity of the region derived from the visual image and the average region temperature derived from the thermal image. Also, a histogram of the values of W_{cd}/W_{abs} for each region is computed, and the mode of the distribution is chosen to represent the heat flux ratio for that region.

The classification of regions is based on rules which use the above features. The rules are of the form:

IF {VALUE(R) e [0.2,0.9] AND VALUE(reflectivity) e [0.35,1.0]}
OR {VALUE(R) e [-.8,-.3]} THEN IDENTITITY = BLDNG

Rules of the above form were derived for each class of object to be identified. The intervals were specified heuristically based on observed vari-

ations in the values among different regions of the same class.These rules were encoded in a decision tree classifier.

We tested this approach using real data gathered from naturally occurring outdoor scenes [4]. Figure 3 shows the visual image of a scene imaged at 1:30 pm in March. Figure 4 shows the thermal image of the same scene. A histogram of values of the ratio W_{cd}/W_{abs} is computed for each region, and the mode of each distribution is obtained (Figure 5). The surface reflectivity (Figure 6) of each region and the average region temperature (Figure 7) are also computed. These features are used by the classification algorithm discussed above, which labels to each region as a vehicle, building, vegetation, or road, as shown in Figure 8.

The method described above was tested on other similar sets of data obtained at different times of the year and obtained results consistent with those presented here. Figures 9, 10 and 11 show data and results acquired from a tank surrounded by vegetation. Again, the values of the heat flux ratio for vegetation are highest, while those for the tank are lower. The classifier used for the previous experiments, however, will fail for this object since the lumped thermal capacitance of the tank is much higher than that of automobiles. The classifier, therefore, must be designed for the domain of application. In other words, rules for recognition are task-specific.

The results discussed above were presented at the First International Conference on Computer Vision (London, 1987) [5]. A paper based on these results was also published in the *IEEE Transactions on Pattern Analysis and Machine Intelligence* [4]. Generalization of these results was presented at the IEEE International Conference on Robotics and Automation [2]. Our group has presented these results at the NATO Advanced Research Workshop on Highly Redundant Sensing for Robotic Systems [6] and the NATO Advanced Research Workshop on Multisensor Fusion for Computer Vision [7], and at the NSF Workshop on Range Image Processing [8].

The work described above was extended to the integrated modelling of thermal and visual image generation. Preliminary results based upon the integrated modeling of thermal and visual images were presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (San Diego, 1989) [9]. Figures 12 - 14 show results obtained from that work.

A related paper on the simultaneous modeling of three dimensional objects for visual and thermal image synthesis was presented at the 1990 Optical Engineering Southcentral Symposium [10]. A volume surface octree is used for object modeling. This representation was found to be suitable for thermal modeling of complex objects with non-homogeneities and heat generation. The technique to incorporate non-homogeneities and heat generation using octree intersection was discussed. The proposed model can be used to predict discriminatory features for object recognition based on the surface temperature and intrinsic thermal properties in any desired ambient condition. The model is designed to be used in a multisensor vision system using a hypothesize and verify approach. Several examples of the generated thermal and visual images were presented to illustrate the usefulness of the approach.

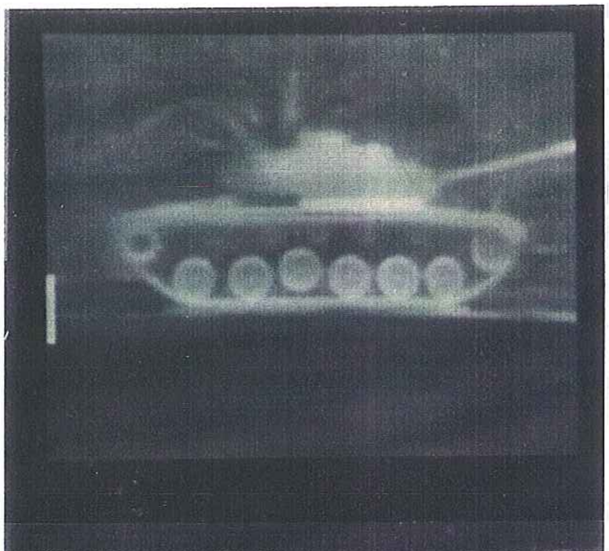

Figure 9. Thermal Image

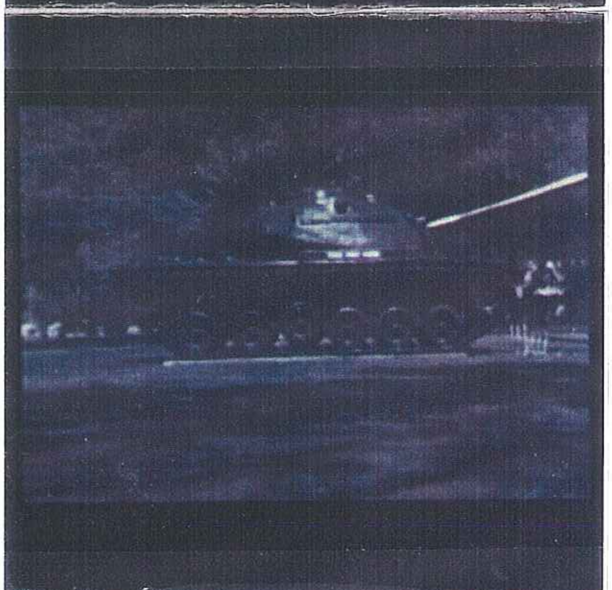

Figure 10. Visual Image

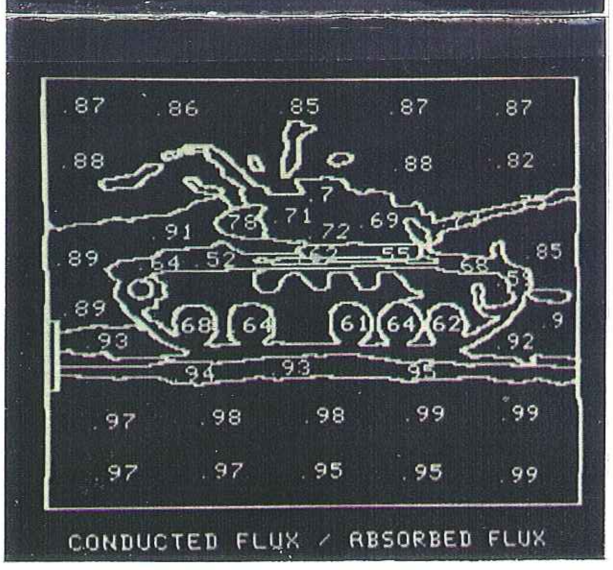

Figure 11. Ratio of conducted heat flux to absorbed heat flux

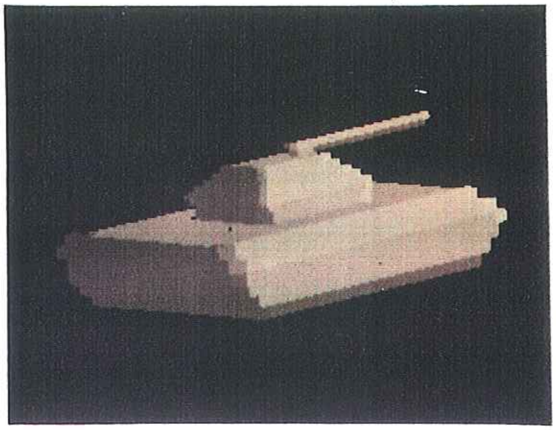

Figure 12. Visual Image of Object.

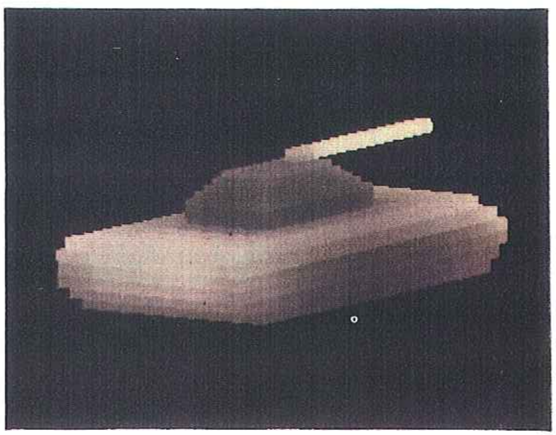

Figure 13. Temperature image of Object. Max. temp. = 316K Min. temp. = 303K

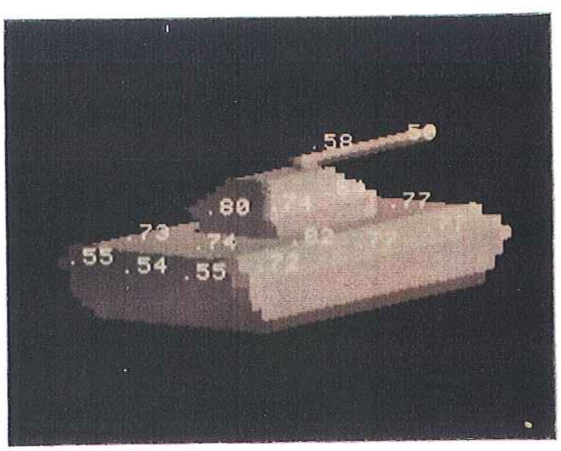

Figure 14. Distribution of Surfaces Heat Flux Ratio: (Conduction/ Absorbed) Mode = 0.75

2. SEGMENTATION AND UNDERSTANDING OF LADAR IMAGES.

Continuing our efforts at understanding multisensor images, we have developed a system for image segmentation, analysis, and understanding using laser radar and thermal images. This work includes several new segmentation techniques for multisensor images and a prototype expert system for interpreting these segmented results, which are briefly described below.

We have studied ladar images of manmade objects in outdoor scenes with the objective of separating man-made objects from the background. At first, we explored ways to integrate information from two sources and analyzed results using both ladar range and ladar intensity data to improve segmentation. We used planar surface fitting to segment the range image. The background usually could not be fit into planar segments. Objects, on the other hand, yielded planar segments. The intensity image was segmented by finding statistics of local busy-ness. We intersected these segmentation maps to generate a combined result. The integrated segmentation results showed strong resemblance to human-generated segmentation, and shared nearly coincidental region contours. Further, by examining the computed means and standard deviations, we found that (*i*) different types of targets generate different statistics in intensity data, and (*ii*) the background segments have much higher standard deviations in range data than object segments. These preliminary results, based on the segmentation of range and intensity portions of images, were presented at the 1988 Conference on Pattern Recognition for Advanced Missile Systems [11]. We continued the development of this system with the addition of velocity images of the ladar signal. Results based on velocity, range, and intensity were presented at the Sensor Fusion Workshop II: Human and Machine Strategies (Philadelphia, 1989) [12], and published in *Pattern Recognition* [13]. In addition, we have added a module for segmenting thermal images. The segmentation obtained from the four modalities is combined using a procedure described later in this section.

We have developed a prototype system for interpreting segmented laser radar (ladar) images for man-made object recognition and image interpretation. The objective of this prototype system is to recognize military vehicles in rural scenes. The system uses a knowledge-based system which is constructed using KEE rules and LISP functions, and uses results from preprocessing modules for image segmentation and integration of segmentation maps. Low-level attributes of segments are computed and converted to KEE format as part of the databases. The interpretation modules detect man-made objects from the background using low-level attributes. Segments are first grouped into objects, and then man-made objects and background segments are classified into pre-defined categories, such as tanks, ground, etc. A concurrent server program is used to enhance the performance of the knowledge-based system by serving numerical and graphical-oriented tasks for the interpretation modules [12]. Complete results on this expert system will be presented at 1991 Conference on Artificial Intelligence Applications [14], and will be published in *Machine Vision and Applications* [15].

In addition, we have developed a new approach for segmenting scenes using multisensor data based on the pyramid data structures. In this approach, image pyramids are built for each sensing modality. Information fusion between these pyramids is used to establish the scene segmentation. We applied the technique to real multisensor data to test its performance. The segmentation which results from this technique is suitable for use by vision systems which classify objects (image regions) using multisensor data. A paper containing the details of our approach and experimental results was recently published in

Pattern Recognition [16]. The application of this approach to thermal and visual scenes, and the advantages of this approach over previous techniques which use a single imaging modality are also discussed.

Recently, we have formulated and implemented an interesting method for combining region and edge-based segmentation. These results were presented at the 1990 International Conference on Computer Vision in Osaka, Japan [17]. This algorithm integrates segmentation maps using both region and edge segmentation maps as input to obtain a region map in which each region is large and compact. The operation is efficient and independent of image sources as well as segmentation techniques. The algorithm allows multiple output maps and applies user-selected weights on various information sources. The scope of integration is parametrically controlled for the desired spatial resolution. A maximum likelihood estimator provides initial solutions of edge positions and strengths from multiple inputs. An iterative procedure is then used to smooth the resultant edge patterns. The edge map is converted to a region map, using closed edge contours if desired. Finally, regions are merged to ensure that every region has the required properties. Experimental results are demonstrated using various segmentation techniques and real data from laser radar and thermal sensors.

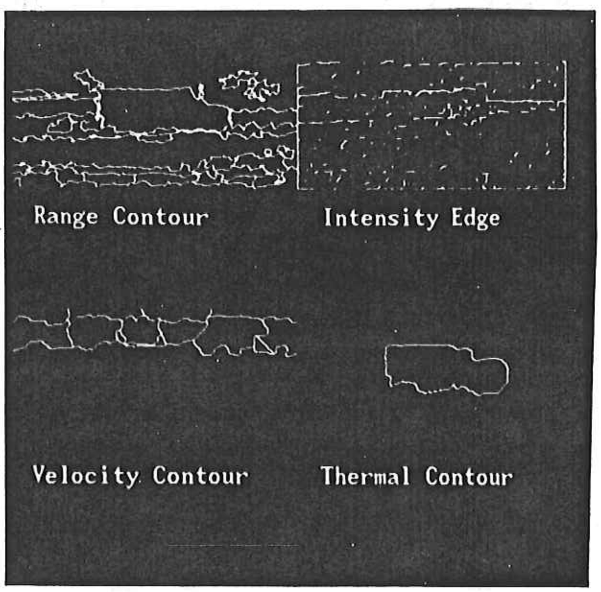

Figure 15. Four input region or edge maps

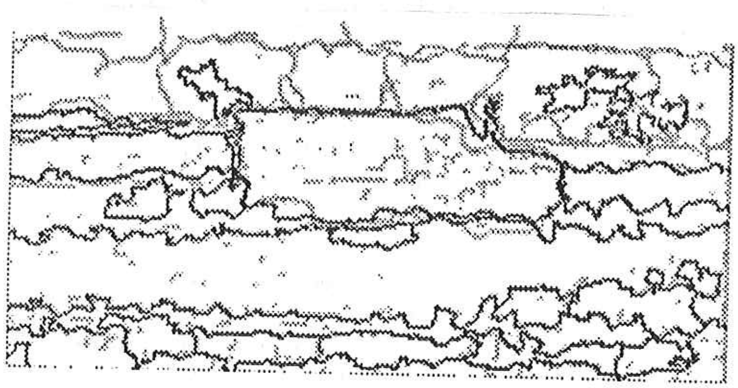

Figure 16. An example of the integrated segmentation produced by the system.

3. INTERPRETATION OF RANGE IMAGERY.

The Computer and Vision Research Center is a leader in the interpretation and understanding of range images. One of the earliest contributions on this subject was presented by Vemuri and Aggarwal at the 1984 International Conference on Pattern Recognition [18]. Subsequent contributions were published in the journal, *Image and Vision Computing* [19], in the book, *Three Dimensional Machine Vision* [20], in *Proc. Conf. on Computer Vision and Pattern Recognition* (1986) [21], and in *IEEE Trans. Circuits and Systems* [22].

Over the past five years, active devices have been developed which can provide three-dimensional data directly to vision systems. We have examined the progress that has been made in the field of 3-dimensional computer vision, from data acquisition to object recognition, and reviewed published research results [23]. This review of the state-of-the-art of 3-D computer vision was presented at the 1988 International Conference on Pattern Recognition [24].

As described in greater detail below, we have developed a hierarchical approach for segmentation of dense, 3-D range images. Our first attempt at segmentation used four local properties (the 3-D coordinate, the surface normal, the Gaussian curvature, and the mean curvature of each data point), combined in a hierarchical data structure to segment a given 3-D dense range map into surface patches. This algorithm applies to planar and curved surfaces. This research was presented at the 1988 IEEE Conference on Systems, Man, and Cybernetics [25]. Subsequently, we developed a more robust algorithm based upon the surface normal and its projections, which we presented at the recent International Conference on Computer Vision in Osaka [26].

A model-based vision system has been developed in which a commercial CAD system is used for object modeling. Assuming that the model is known, the corresponding object in the scene is located. Given the CAD model of an object, certain features of the model are extracted, while others are precalculated and stored. Using the newly developed segmentation procedure, the given dense 3-D range image is segmented into a set of homogeneous surface patches. Properties such as curvature, surface normal, and surface area are approximated for each surface patch. For each extracted surface patch, three filters are applied to the previously obtained model features to find the best match. Then a global consistency filter is applied to remove ambiguities and to find the best matched model [27]. In addition, we have developed methodology for constructing octrees from range images [28].

In a study on the determination of motion from a sequence of range images, we have developed an algorithm that uses extracted planar patches from the scene to estimate motion. Given the correspondence between planar surface patches in a sequence of 3-D range images, and assuming that the object is rigid, the motion transformation is estimated. The plane surface parameters are used to formulate a least square optimization problem that computes the optimal rotation and translation. This results in the transformation that best fits the images. This algorithm proved to be reliable on both synthetic and real data [29].

Another aspect of our research relating to object recognition, in which occluding contours are used for model construction and shape recognition, is given in a paper by Chien and Aggarwal, published in the *IEEE Transactions on Pattern Recognition and Machine Intelligence* [30]. This approach, which is based upon octree descriptions of each model and a hypothesis/verification process, allows planar and curved objects to be handled in a uniform manner.

The octree structure is a popular means for representing the volume of 3-D objects. When surface information is encoded into an octree, the resulting octree is called a volume/surface octree (VS octree). The VS octree is a compact and informative representation of 3-D objects.

In the following, we present partial details some of our most recent research contributions in range image understanding. The work described in the section entitled "Segmentation of Range Images," was presented at the 1990 International Conference on Computer Vision (Osaka, Japan) [26]. The work described in the section entitled "Motion from a Sequence of Range Images" was presented at the 1990 International Workshop on Intelligent Control (Istanbul, Turkey) [29].

- Segmentation of Range Images.

The first step in any object recognition task is to partition the input data into homogeneous regions and extract a set of primitives. In this section we address the problem of partitioning range imaging that represent the 3-D coordinates of each point in the scene. Specifically, the problem addressed is stated as follows: *Given a 3-D range image of a scene containing multiple arbitrarily shaped objects, segment the scene into homogeneous surface patches.*

Much work has been done in the past on the segmentation of range images. Besl and Jain [31], Fan, Medioni, and Nevatia [32], and Brady, et al. [33], segment the images by using the Gaussian curvature and the mean curvature to determine the similarities and dissimilarities in the data. Hoffman and Jain [34] cluster the input data into regions by detecting connected planar, convex, or concave surfaces using various statistical measurements made on several properties of surfaces. Most of the remaining range segmentation approaches depend upon the types of expected surface shapes. For example, Boulanger and Rioux [35] segment planes, spheres, ellipsoids, and other simple quadric surfaces; Han, Volz, and Mudge [36] and Flynn and Jain [37] segment planes, spheres, and cylinders. While these approaches may be well suited for specific applications, they lack the generality needed for most real world applications. The common drawback to most algorithms is their requirement for many empirically determined thresholds. Since these thresholds are dependent on the quality and the type of the input data, such algorithms can only be used for a small class of input data. For each class of input data, the thresholds must be readjusted. For the algorithm to be independent of its input data, it is important that the thresholds be derived from the data itself.

In general, a segmentation procedure partitions a given image into homogeneous regions. The segmentation should depend on the input data type and on the final representation of the homogeneous regions. These qualifications suggest a two-part framework for segmentation: one driven by the input data and a second driven by the final region representation. In this procedure, we adopt such a modular framework. Figure 17 shows the overall organization of the segmentation scheme. The procedure is divided into two modules. The first module is the low level segmentation module where the local properties are extracted from the given input data, and clustered into homogeneous regions. This module gives an initial over-segmented output. These over-segmented regions are merged in the second module using the surface representations dictated by higher vision tasks.

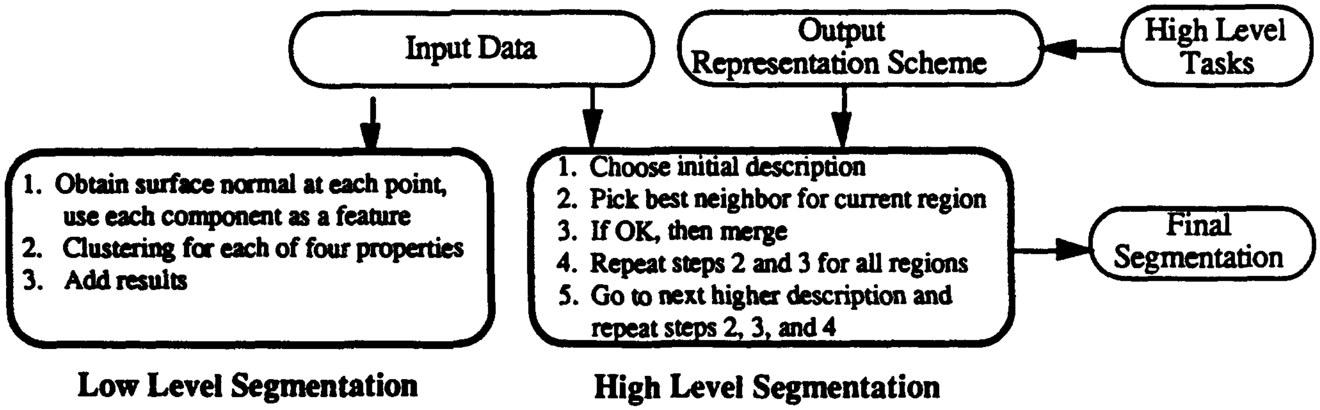

Figure 17. Overall Organization of the Segmentation Scheme

Low Level Segmentation Module. The low level segmentation module uses local information to arrive at a preliminary segmentation of the input data. The module is divided into two stages. The **preprocessing stage** computes the local properties of the input data, and the **pyramidal clustering stage** clusters the data into homogeneous regions. The clustering is performed using four properties which are calculated by the preprocessing stage for each point in the range image. The four properties are the surface normal vector, its three projections onto the xy-plane, the yz-plane, and the xz-plane. The projections are equivalent to the views generated by viewing the scene from three orthogonal directions. Prior to calculation of these properties, smoothing is performed to reduce the noise level. Each of the four images generated by the preprocessing stage are independently used by **pyramidal algorithms**, resulting in four initial segmentations. The pyramidal algorithm is an iterative procedure that clusters pixels with similar properties into groups in a hierarchical manner. The four segmentation outputs are then added, resulting in a maximally partitioned image. The result of this module is an over-segmented image.

High Level Segmentation Module. The resulting segmentations from the first stage represent a local grouping of the corresponding local property. The second stage of the procedure merges the regions based on certain homogeneity criteria, which depend on the final representations of the surface patches to be used by the high level tasks. This stage can be modified according to the application that uses the segmentation results. Here we use bivariate polynomials of up to fifth degree to represent each patch. Two adjacent surface patches are merged if parameters of one of the patches results only in a small error when used to extrapolate over the neighboring patch.

Figures 18-22 show results obtained using this segmentation procedure. Range images were obtained from several institutions and the parameters were unchanged for all the examples shown. The details of the segmentation procedure are found in the paper by B. Sabata et al. [26]. The original range image is shown on the left, and the results on the right.

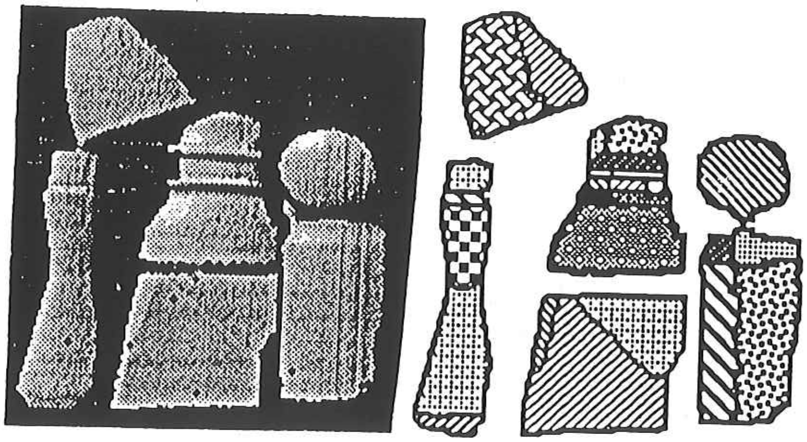

Figure 18. Left: Complex range image of several objects obtained by 100A Technical Arts Laser Scanner (The University of Texas at Austin). Right: Segmentation results.

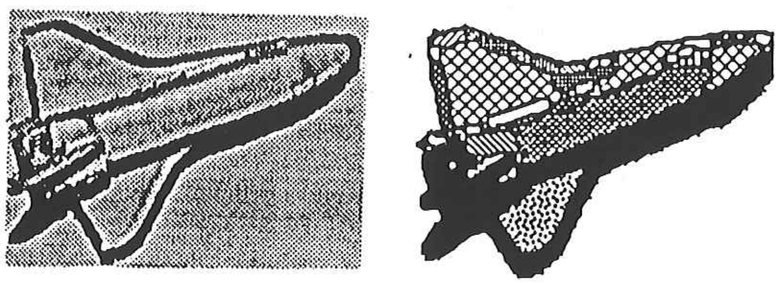

Figure 19. Left: Range image of model space shuttle (obtained from SRI International). Right: Segmentation results.

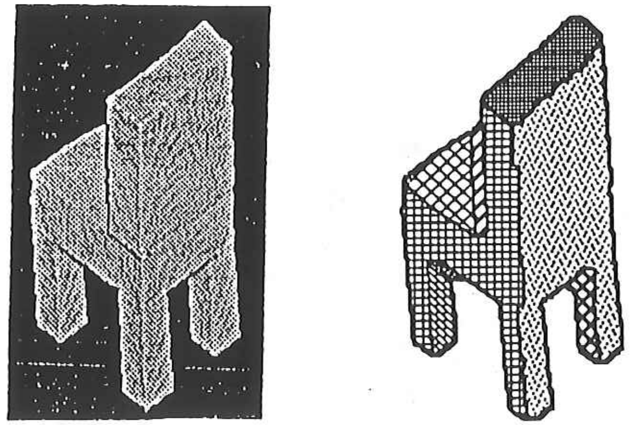

Figure 20. Left: Range image of chair (obtained from University of Southern California). Right: Segmentation results.

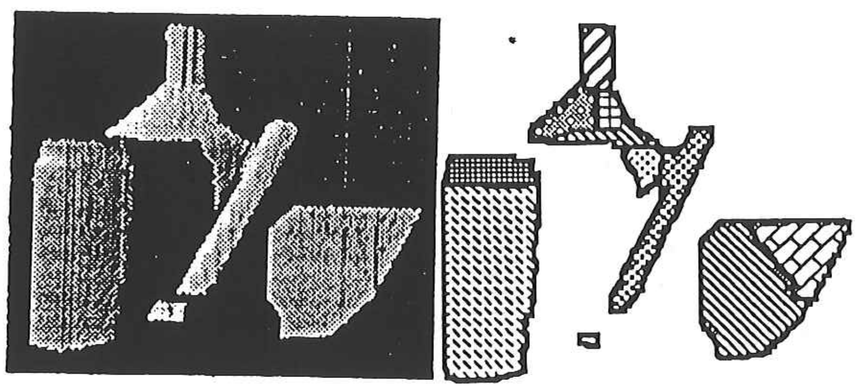

Figure 21. Left: Complex range image of several objects obtained by 100A Technical Arts Laser Scanner (The University of Texas at Austin). Right: Segmentation results.

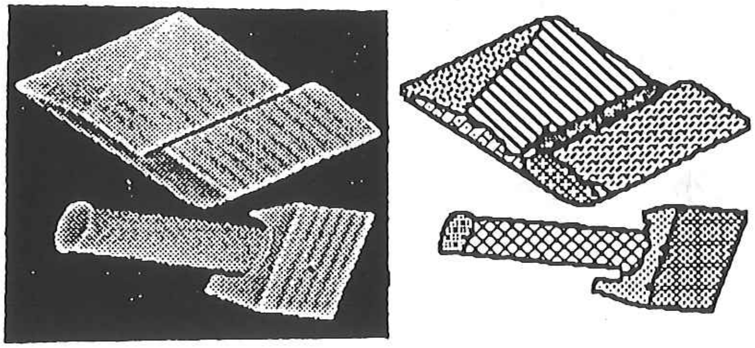

Figure 22. Left: Image of scene containing two polyhedrals, one atop the other, and a third object with a cylindrical region (obtained from Michigan State University.) Right: Segmentation results.

• <u>Motion from a Sequence of Range Images.</u>

A new robust method to estimate motion from a sequence of range images has been developed, which uses the correspondence between planar surfaces in a sequence of images rather than point and line correspondences. Rotation and translation are both determined using properties of the planar surfaces, such as surface normal and surface intersection. Because global features are used, the method is less sensitive to noise, quantization errors, and partial occlusion of the surfaces. This method can be applied to a scene containing several objects,each with a different motion. The algorithm has been tested on sequences of synthetic data as well as laser range data.

Many algorithms have been developed for computing motion from sequences of stereo images which rely on point feature correspondence. These algorithms are very sensitive to noise, particularly errors which accrue due to quantization of disparity. Very few investigators have studied estimation of motion from a sequence of range images. Lin, et al., [38] and Aloimonos and Rigoutsos [39] discussed correspondence-less methods of estimating motion from sequences of images. Their main assumption is that a point feature visible in one image must be visible in the next. Correspondence is not required. They also assume that the set of points lie on a plane. The assumption that the same points are visible from one image to the next generally does not hold for real data acquired from available laser ranging systems.

In most existing algorithms, point or line correspondences in the sequence of images are established prior to estimating motion. In practical situations, occlusion and noise greatly complicate the establishment of point or line correspondence. Unlike 3-D points, a surface in that appears in one image is more likely to appear in the next. It is unlikely that the entire surface will be completely occluded. Hence, using surface correspondences could minimize problems due to occlusion. The sensitivity to error of individual features does not argue against information in the aggregates; however, since the surface is derived from a large set of points, as compared to line and point features, the surface fitting process suppresses the contribution of noise and quantization errors in individual points.

No previous research has been reported using surface correspondences to compute the motion of objects. Unlike stereo data, laser range data are dense and complete surface information is available (not just the discontinuities). Thus, sequences of laser range data are more suitable for motion estimation using surface correspondences. Using the assumption that surface correspondences have been established *a priori*, we have developed a robust algorithm for estimating motion of objects containing planar surfaces. This algorithm was presented at the Image Understanding Topical Meeting [40] and published in the *Proceedings of the IEEE International Workshop on Intelligent Motion Control* [29].

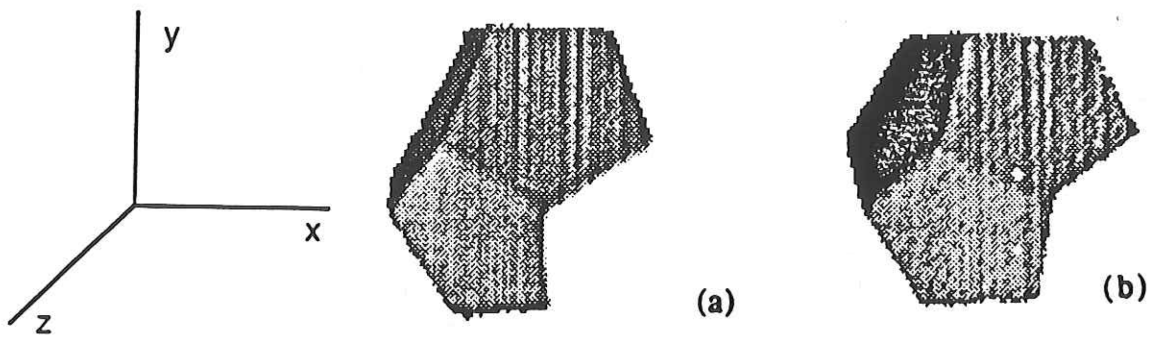

Figure 23. Real image sequence acquired from a laser range scanner.

The algorithm was applied to a sequence of range images. Figure 23 shows a sequence of real range images acquired with a 2000A White Scanner. The scanner output is in the form of (x,y,z) coordinates of each point that has been scanned. The scene consisted of a polyhedral object whose three surfaces are visible in both frames. The second frame was obtained by rotating the object by 18 degrees, about an axis parallel to the Y axis, and passing through (0,0,-3).

Results obtained using the motion estimation techniques developed in this paper are given in Table 2. The motion is expressed in the form of a rotation about the X, Y, and Z directions. In this example, a real image is noisy and the third surface patch in the frame is barely visible. The segmentations results give just 30 points belonging to the patch. In spite of this, the algorithm performs well and estimates are close to the ground truth.

		Actual	Estimated
Rotation	θ_x	0°	0.75°
	θ_y	18.0°	15.8°
	θ_z	0.0°	1.55°
Translation	X	0.927 in.	0.993 in.
	Y	0.0 in.	0.14 in.
	Z	0.147 in	0.103 in.

Table 2. Motion estimation for real data. Estimated motion is in the form of a rotation about the coordinate axes and then a translation.

Related Motion Research.

In the course of pursuing research on the computation of motion, we compiled an overview and comparative study of the literature on the computation of motion, which was published as an invited paper in the *Proceedings of the IEEE* [41].

In a related study, we developed a two-stage solution to the problem of correspondence of points for motion estimation in computer vision. The first stage of the algorithm is a sequential forward-searching algorithm (FSA), which extends all the survivor trajectories. The second stage of the algorithm is a batch-type, rule-based, backward-correcting algorithm (BCA). Under the simple error assumption (no chain errors), seven rules are sufficient to handle all the possible errors made by FSA. BCA takes the last four frames of points as input and rearranges the correspondence among them according to these rules. FSA and BCA are applied alternately. This algorithm is able to establish the correspondence of a sequence of frames of points without assuming that the number of points in all frames are equal or that the correspondence of the first two frames has been established. Experiments have illustrated the robustness of the algorithm on sequences of synthetic data as well as on real images. [42]

Finally, we have addressed the problem of reconstructing a 3-D line from a sequence of monocular images (2-D projections), in a paper presented at the 1990 International Conference on Computer Vision [43]. In this paper, we first consider the problem of 3-D line representation and then the recursive estimation algorithm. We point out the problems with all previous 3D line representation models, and suggest a new approach based on simple geometrical observations. We then derive the corresponding recursive estimation algorithm for the new representation, based on simple linear algebra. Simulation results from implementing the representation model and the recursive algorithm on an IBM RT PC are presented.

4. PASSIVE AERIAL NAVIGATION BY IMAGE SEQUENCE ANALYSIS

With the advent of sophisticated techniques for sensing electromagnetic emissions, passive navigation of aircraft has become of vital importance to the military community. Passive navigation is the determination of one's position and heading without the emission of electromagnetic radiation by the aircraft. Active navigation, on the other hand, must rely on such emissions. Because the electromagnetic emissions can be detected, active navigation is unsuitable for many military applications. Therefore, there is a serious need for passive navigation systems.

In this research, we are developing a passive navigation system based on matching a sequence of aerial images to a digital elevation map. Specifically, the research problem is stated as follows: *Given a sequence of aerial images and a digital elevation map, determine the trajectory of the aircraft relative to the digital elevation map.*

Digital elevation maps (DEMs) have recently become more practical for use in aircraft avionics systems. (A DEM is a digital database of uniformly spaced terrain elevation measurements.) Recent advances in computer performance and mass storage technology have led to increased interest in using DEMs for navigation and other applications. In particular, DEMs are attractive for use in aircraft navigation systems.

In our research, a general navigation system is being developed based upon computer vision techniques. Since it is desirable to determine one's position without resorting to active emissions, our system uses a sequence of aerial images as the primary input. A DEM is used as the reference database for feature matching. (Presumably this would be carried on-board the aircraft.) This research has produced encouraging results, which have been presented at several workshops [44], [45] and the 1990 International Conference on Computer Vision [46]. Our most recent results were published in *IEEE Transactions on Pattern Analysis and Machine Intelligence* [47].

A DEM of a region in Colorado was obtained from the United States Geological Survey. The lack of available real aerial image sequences prompted us to simulate such images using the DEM. This required combining of a variety of techniques. Lambertian shading was used to compute intensity values corresponding to the DEM samples. Next, an arbitrary aircraft trajectory was chosen and perspective projection was used to generate the perspective aerial image sequence. Figures 24 and 25 show the first two images in this sequence. In a real situation, these images would be acquired from an aircraft.

A stereo algorithm was implemented to recover estimated elevation maps from the intensity image sequence. The first step in this procedure generates a set of edge maps using several scales of resolution for each image. Two successive images were treated as the left and right stereo images. For example, the image in Figure 24 is treated as the left image, and the image in Figure 25 is treated as the right image. The edge points in their corresponding edge maps are matched to create a disparity map. (The disparity of an edge point is the amount of displacement between the edge point's x coordinate in the left image and its x coordinate in the right image.) From this disparity map, an estimated elevation map is reconstructed by applying the inverse perspective projection equations.

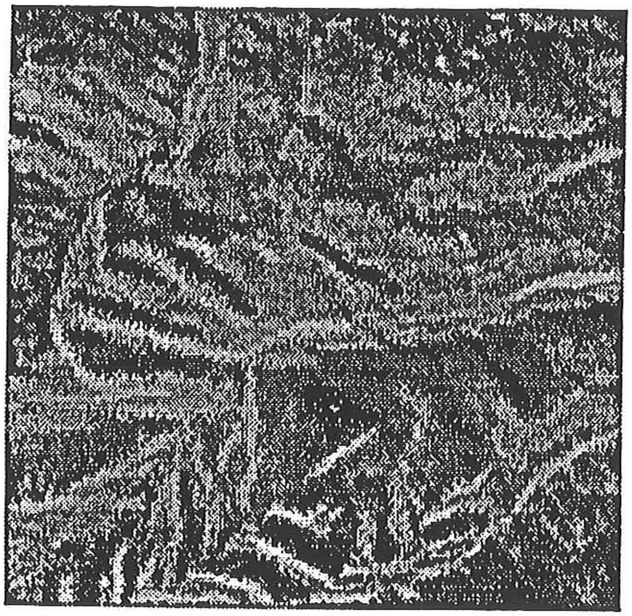

Figure 24. Left Image

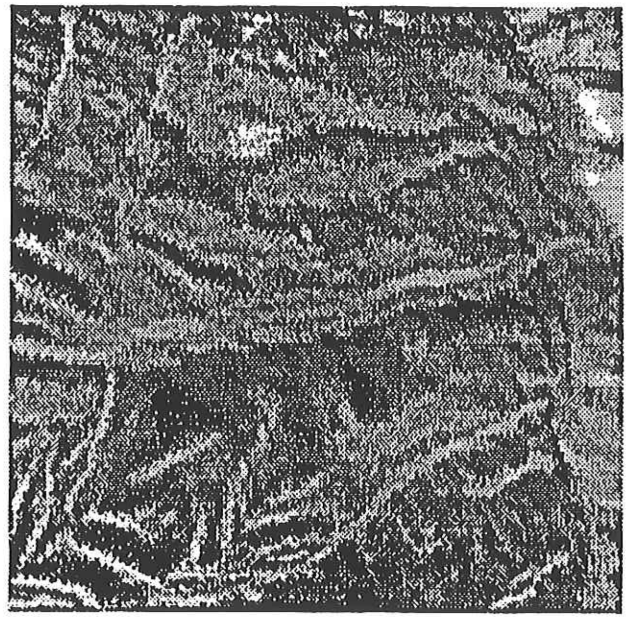

Figure 25. Right Image.

From the estimated elevation maps, the significant three-dimensional features are extracted -- namely, valleys and ridge lines. Determining the position and orientation of the recovered elevation map (REM) within the reference DEM requires a 3-D surface matching algorithm. Because of the sparseness of the disparity map and the smoothness of the resulting interpolated surface, a meaningful 3-D curvature measurement cannot be accurately computed. For this reason, we developed a novel terrain representation--called a *cliff map*--that is computed for both the REM and the DEM. For matching, the REM and the DEM are converted to the cliff map representation. Cliffs are defined to be the zero-crossings obtained after convolving the elevation map with a Laplacian-of-Gaussian (LoG) filter. Edge detection is applied to an elevation map in the same manner as it is traditionally applied to optical intensity images. The cliff map representation is computed for both the recovered elevation map and for the reference digital elevation map. This computation in effect transforms the 3-D matching problem into a 2-D matching problem. The cliffs in the REM must be matched to those in the DEM to determine the position and orientation of the unknown image. Rather than attempting to match the entire cliff contours, critical points are extracted to form a more compact representation. These critical points then serve as the basis for a point-based matching algorithm. The feature map derived from the images in Figures 24 and 25 is shown in Figure 26. In this figure, the two darkest gray levels represent two levels of valleys and the two brightest gray shades represent two levels of ridge lines. The same feature extraction techniques are then applied to the DEM to create the feature database to which the REM must be matched. Figure 27 presents this digital elevation map, and Figure 28 shows the REM overlaid onto the DEM.

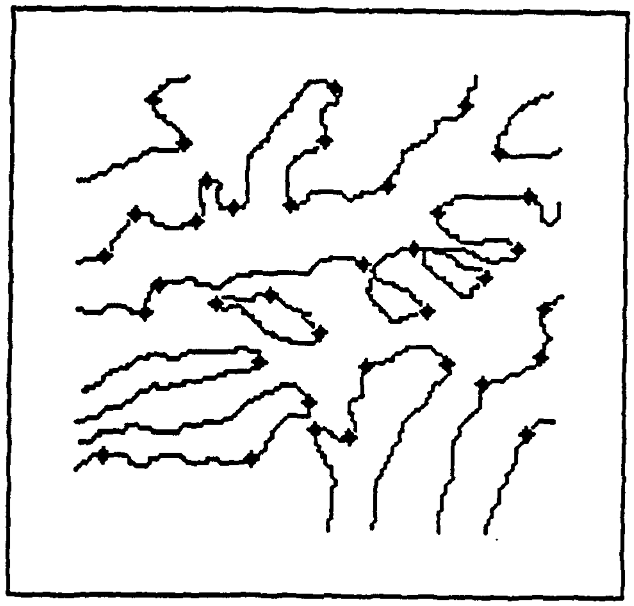

Figure 26. REM Cliffs and Critical Points

As seen from the overlay in Figure 28, the technique of matching cliff maps, recovered elevation maps, and given digital elevation maps yielded excellent matching. These results are based upon cliff maps that were recently developed by our group [44], and earlier results of other researchers [48]-[50] and of our group [51]-[53]. In this test it was not possible to use real aerial images.

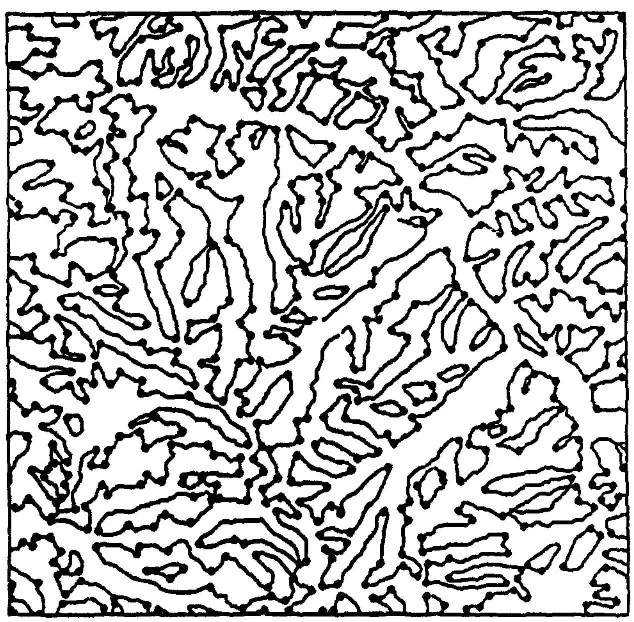

Figure 27. DEM Cliffs and Critical Points

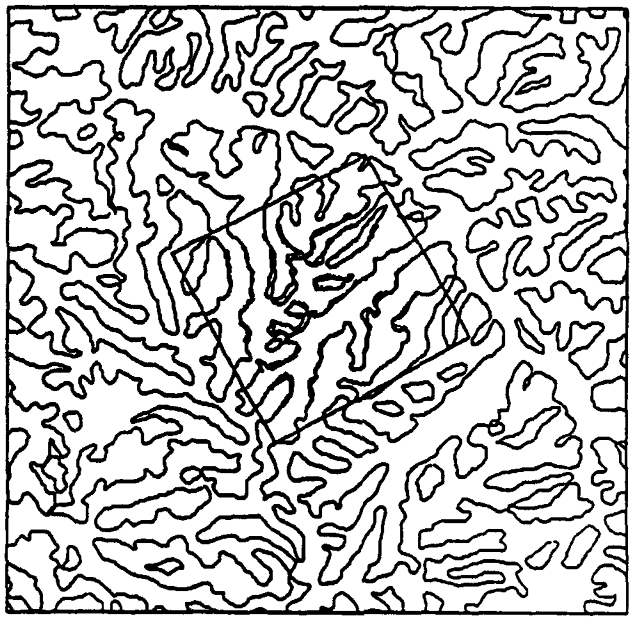

Figure 28. REM Overlaid onto DEM

By matching optical intensity images to 3-D terrain data, a continuous report of an aircraft's position and heading can be obtained. Processing of the reference digital elevation map data can be performed offline and carried onboard the aircraft. Use of the *cliff map* representation provides a compact representation of the terrain, and use of critical points in the matching strategy makes testing of every possible position and orientation unnecessary. Furthermore, the matching algorithm is well-suited for parallel implementation since each hypothesized match can be verified independently. Experiments were performed using real terrain data to assess the robustness of the terrain matcher. It was found that a successful terrain matching could be achieved even when the stereo match rate is 70% and the disparity value error rate is 40%. These results indicate that the use of cliff map contours for terrain matching is both efficient and robust.

The results of this research may be applicable to several areas. Not only is this work important for navigation of aircraft, but for navigation of other vehicles as well. For example, there is a similar research effort in progress for autonomous land vehicles. As with aircraft, the stealth of autonomous land vehicles also depends on passive navigation. Another application is the registration of aerial reconnaissance images. By matching a previously-obtained aerial image sequence with a DEM, the precise location of the area can be determined. This research is therefore expected to have a great impact in many areas where passive navigation is necessary or desirable.

Our study of digital elevation maps has led to some interesting related research on stereo imaging systems and algorithms. In designing a stereo imaging system, one must consider how the various system parameters affect the range estimation error. We conducted a stochastic analysis of the quantization error in a stereo imaging system. The probability density function of the range estimation error and the expected value of the range error magnitudes were derived in terms of the various design parameters. We found that when the depths in the scene lie within a narrow range, a better

measure of range resolution is obtained by the use of the *relative range error* ($\varepsilon = \frac{|\Delta z|}{z_{max} - z_{min}}$ rather than the percent range error ($|\Delta z|/z$). These results were presented at the Conference on Computer Vision and Pattern Recognition [53] and published in the *IEEE Transactions on Pattern Analysis and and Machine Intelligence*.[54].

In another thrust of our continuing research in this area, we evaluated the contribution of a third camera to increasing accuracy in stereo correspondence [55],[56]. The use of a third image requires additional computations for preprocessing the extra image. Understanding the trade-off between the contribution of the third camera and the additional computation required to use the third image is of paramount importance in the design of real-time stereo based vision systems. We evaluated the relative performance of binocular and trinocular stereo algorithms on aerial monochrome images generated from digital elevation maps in order to obtain accurate ground truth verification. We developed a new methodology for comparing the matching performance of stereo algorithms using actual digital elevation maps. From these experiments, we found that trinocular local matching reduces the percentage of mismatches with large disparity errors by more than one-half, as compared to binocular matching, while increasing the computational cost of local matching by approximately one fourth. These results were presented at the *Image Understanding and Machine Vision Workshop* [55] and the *IEEE International Conference on Robotics and Automation* [56] and will soon be published in the *International Journal of Computer Vision* [57].

5. IDENTIFYING MAN-MADE OBJECTS IN OUTDOOR SCENES / FUSION OF COLOR AND GEOMETRY INFORMATION

In past research, we investigated the use of color information to interpret images of outdoor scenes (especially aerial images) in great detail. We successfully applied these techniques to the segmentation and classification of regions in aerial photography [58]-[61] and devised segmentation of more general classes of chromatic images [62]. In our research efforts under this contract, we developed a new approach for the detection and segmentation of man-made objects in color images of natural scenes [63], [64].

The approach is based on detecting geometric structure in the image and combining the detected structure with color information to guide the segmentation. The central problem is the detection and semantic interpretation of large stationary man-made objects in color images of nonurban scenes. We describe the generation of color-based confidence functions for material selection in incremental segmentation. We focus on the segmentation of images of concrete bridges. These techniques are applicable to autonomous navigation, target navigation, target acquisition, and several industrial computer vision problems. Large concrete objects often have rectilinear edge structures with many parallel relationships. We use these properties to guide our initial incremental segmentation toward concrete objects. The goal for our segmenter is to locate the representative faces of concrete material in the image as a starting point for the interpretation phase. These heuristics rely on the detection of straight lines segments of the edge map of the gray-scale image. The straight line segments, once detected, are then grouped according to several perceptual grouping criteria. The straight line segments are then constrained further by region label restrictions. Finally, color cues are used to restrict the candidate artifacts further and to produce confidence measures of our initial belief in our estimation of the material of the identified rectilinear faces [64].

The results of this processing are fed into an expert system. Truth maintenance techniques are used to reason about all the candidate artifacts and decide which ones correspond to the model of a known object. The expert system shell is able to request further processing from the low level algorithms to clear ambiguities. Finally, the expert system displays its interpretation of the scene. This research was presented at the 1988 SPIE Conference [63], the IEEE Computer Society Workshop on Computer Vision [65], and documented in the archival *Journal of the Optical Society of America* [64]. Although this approach is based on a two-dimensional model of the side view of the bridge, the recognition technique also works with non-optimal (non-orthogonal) viewing angles.

We have extended this approach to 3-D hypothesized representations of the world, projected to 2-D for tentative match with the observed image. The central problem was to determine what perspective projection parameters should be used to derive a 2-D hypothesis from the 3-D model. As it is practical in most real world situations, we assumed that the geometry of the camera was known, as well as its roll and pitch. We used the fact that many large, man-made structures have prominent straight lines in known 3-D directions, such as vertical and horizontal. In a perspective projection, parallel straight lines in 3-D converge to vanishing points, the location of which depends upon projection angles. By detecting likely vanishing points in the image, under geometric constraints, we could derive projection parameters. Furthermore, we could classify observed straight lines according to their most likely 3-D

orientation for matching with the model. This method was demonstrated on images of various bridges. These results were presented at the 1989 Scandanavian Conference on Image Analysis [66] and at the NATO Advanced Research Workshop on Multisensor Fusion for Computer Vision [67].

6. POSITIONAL ESTIMATION TECHNIQUES FOR AN OUTDOOR MOBILE ROBOT

The autonomous navigation of mobile robots is an important area of computer vision research. Before a mobile robot can perform any useful task it must have the ability to estimate its position and pose accurately in the environment. Many techniques have been suggested for solving this problem of self-location. Using the wheel encoders provided on the mobile robot for positional estimation is not very reliable, since the information from these encoders is differential. Due to wheel slippage and quantization effects, these estimates of the robot's position contain small errors which accumulate quickly as the robot moves, and the position estimate becomes increasingly erroneous. A popular technique is to aid the robot in the positional estimation process by providing alternate means of sensing the environment by using visual and/or range sensing devices.

Various techniques have been studied for estimating the position and pose of an autonomous mobile robot using different kinds of sensors. The techniques vary, depending on the kind of environment in which the robot navigates, the known conditions of the environment, and the type of sensors with which the robot is equipped. The position estimation techniques can be broadly classified into the following four types: (1) landmark-based methods; (2) methods using trajectory integration and dead reckoning; (3) methods using a standard reference pattern; and (4) methods using a priori knowledge of a world model and matching sensor data with the world model for position estimation.

Our present work on position estimation falls into the fourth category--the robot is aided in its navigation tasks by a preloaded world model which provides a priori information about the environment. The basic idea is to sense the environment using onboard sensors on the robot and then to try to match these sensory observations to the preloaded world model. This process yields an estimate of the robot's position and pose with a reduced uncertainty and then allows the robot to perform other navigational tasks. The problem in such an approach is that the sensor readings and the world model may be in different forms. For instance, given a CAD model of the building and a visual camera, the problem is to match the 3-D descriptions in the CAD model to the 2-D visual images.

In this work, techniques are presented for estimating the position of a mobile robot in an outdoor environment. Two kinds of environments are considered; a mountainous natural terrain and an urban man-made environment consisting of polyhedral buildings. In the case where the robot is navigating in an outdoor natural terrain, a Digital Elevation Map (DEM) of the area is assumed to be given [68], [69]. Also, the robot is assumed to be equipped with a camera that can be panned and tilted, as well as a device to measure the robot's elevation above the datum. The robot is not assumed to have the ability to identify and track and landmarks in the environment. The DEM is a three-dimensional database. It records the terrain elevations for ground positions at regularly spaced intervals. The images recorded by the camera are 2-D intensity images. The problem is to find common features to match the 2-D images to the 3-D DEM. The approach suggested is to use the height and the exact shape of the horizon line (HL) and the known camera geometry of the perspective projection to search in the DEM for the possible camera location. The actual search is a two-step process. The first step is a coarser search that reduces the possible locations to a smaller set using the height of the HL in the image plane in different directions; and the second step refines this estimate using the exact shape of the HL.

From the current robot position, images are taken in the four geographic directions: N,S,E, and W. In generating the four geographic views, the tilt angle of the camera is adjusted until the horizon line is clearly visible in the image. This tilt angle ϕ_i (i = N,S,E,W) is then measured. The approximate height H of the camera above the datum is assumed to be known. The height of HL at the center of the image plane in each of these four images is measured.This can be done using basic image processing techniques. Let this height be h_i (i = N,S,E,W). The reason for using the height of the HL at the center of the image plane is that the DEM is assumed to be gridded in the same directions as those from which the images are taken. So the points that project onto the HL at the center all lie along the same grid line in the DEM.

Using the approximate height of the camera H, the tilt angle ϕ_i, and the HL height h_i in one of the directions, e.g., north, the DEM is searched for possible camera locations. That is, a camera location is hypothesized at each of the DEM grid points and the height of the HL, h_i, is back-projected onto the DEM using the camera geometry to see if any elevation points can project to this height. If any such points exist, the camera location is marked as a possible candidate. However, if any grid point is of a height larger than that estimated by the back-projection, then all the camera locations between the current location and this *tall* point are marked as impossible positions and discarded. This heuristic reduces the search space significantly. Similar heuristics are also used to prune the search space by the camera height H. The results of the search process by using one of the images are thus a sparser set of possible camera locations. These are then considered as the possible set for the next search, which searches among this set with the geometric constraints extracted from the image along another direction. This process is continued by successively applying the constraints in all four directions, and the search refines the possible locations to a small set usually clustered around the actual location.

Stage 2 of the search process is used to further isolate the exact location from the possible ones returned by the stage 1 search. Each of these locations is considered as a possible candidate, and the image that would be seen if the camera were located at that location is generated from the DEM using computer graphics rendering techniques. The HLs from these images are then extracted and correlated against the actual image HL to arrive at a measure of their disparity. The camera location corresponding to the location with the lowest disparity is considered as the best estimate of the location. We can further isolate the exact location by generating the images from the points close to the estimated location and checking to obtain a zero error measure.

The case of mobile robot navigating in a structured, man-made, urban environment consisting of polyhedral buildings is considered in next. The 3-D descriptions of the roof tops of the buildings is assumed to be given. Such a description may be obtained from a pair of stereo aerial images or from the architectural plans of the buildings. The robot uses the camera to image the surroundings, each time adjusting the tilt angle so that the roof top edges are clearly visible in the image plane. Now if a correspondence is established between the 3-D descriptions of these edges and their images the position and pose of the robot can be estimated. However establishing this correspondence is in general not a trivial problem. To alleviate this problem, it is proposed to use the geometric relations between the 3-D descriptions of the roof top edges (model edges) to prune the list of possible correspondences. The viewing plane (the plane in which the robot navigates) is divided into distinct, non-

overlapping regions called the *Edge Visibility Regions* (EVRs). These EVRs essentially capture the geometric relations between the model edges with regard to their visibility from various regions in the viewing plane. Associated with each EVR is a *Visibility List* (VL) which is a list of the model edges that are visible in that EVR; also stored for each edge in the VL of an EVR is the range of orientation angles of the robot for which the edge is visible in this EVR. In this research methods for generating the EVRs given a world model are discussed. An upperbound on the maximum number of EVRs that would be generated for a given model is derived [70]. The use of the EVR representation of the environment of a robot for other navigational tasks like path-planning are outlined.

In this research an algorithm *partition* to partition the xy plane into the required EVRs is developed. In developing the algorithm initially the restrictive case that all the buildings in the environment have flat rooftops that are convex polygons, and are of equal heights is considered. The modifications of the algorithm to the more general case are then discussed. In the restricted case, it is sufficient to consider the projections of the rooftops onto the xy plane in forming the EVRs since the tilt angle ϕ is assumed to be measurable.

The algorithm *partition* that divides the xy plane into the desired EVRs, along with their associated VLs, uses three subprocesses called *split*, *project*, and *merge*. The basic idea of the algorithm is to start with the entire xy-plane as one EVR with a NULL visibility list. Each polygon is considered in turn by extending each of its edges, and the EVRs that are intersected are divided into two new ones. The new EVRs then replace the old one and the VLs of the new EVRs are updated to account for the visibility of this edge by considering the edge to be visible in one half-plane, say left of the edge, and invisible in the other. This is handled by the *split* process. For each new polygon considered, the mutual occlusion of the edges of this polygon with the other existing polygons is handled by forming the *shadow region* of these edges on the other existing polygons. This is handled by the *project* process. The idea is to project each edge of this polygon onto the edges of the other existing polygons, and to determine the area in the plane which lies in the shadow and then delete this edge from the VLs of all the EVRs lying in the shadow region. Efficient ways to compute this shadow region are discussed in [70]. Finally the *merge* process concatenates all the adjacent EVRs with identical VLs into one EVR. After partitioning the xy-plane into EVRs, for each model edge in the VL of an EVR the range of orientations of the robot for which this model edge is visible in the EVR is also computed and stored. An efficient method to compute this range is also described.

In the case of the buildings with rooftops that are not convex, the non-convex polygons representing the projections of the rooftops on to the xy-plane are decomposed into a set of adjacent, component convex polygons. Only decompositions without Steiner points are considered, since the modifications to the existing algorithms in this case are straightforward. (A Steiner point is any point on the boundary of the polygon that is not a vertex.) The extra edges added in the process are considered as *dummy* edges, and their visibility is not marked in the VLs of the EVRs. Hence, the self occlusions of the edges of a non-convex polygon are handled by decomposing the polygon into component convex polygons and dealing with their mutual occlusions by using the *project* process. In the case when all the buildings are not of the same height, it is insufficient to just consider the projections of the rooftops onto the xy-plane alone when forming the shadow regions. The *project* process is modified to consider a z-shadow line also. More details are given in [70]. Note that the case of all the buildings of the same height is actually a special case of

unequal height buildings where the z-shadow line is at infinity. In the case when the rooftop of a building is not flat (planar), it is decomposed into convex planar polygons and each of these is considered as a separate polygon; the partition algorithm is then modified as before, in the non-convex case. Also, the z-shadow lines are drawn to account for these component polygons, which are now convex and flat but of different heights.

An interesting and important question related to using this method is how many EVRs are generated using the *partition* algorithm. If this number is very large, it is impractical to use the method. One might think that the number of distinct EVRs would exponentially with m and n the number of polygons and the number of sides of each polygon, respectively. An upper bound on the maximum number of EVRs that would be generated is derived in [70]. and shows that this is polynomial in m and n, $O(n^2m^4)$. This is a very loose upper bound and only shows that the number of EVRs is polynomial. In practice, however, the number of EVRs generated is much smaller than this.

The uses of the EVR representation in positional estimation and path-planning are also detailed in this research [70].

CONCLUSIONS.

From the results outlined above, it is abundantly clear that our synergetic approach to multisensor fusion for computer vision is an outstanding success. The work has been enthusiastically received by the computer vision community. Our results have been presented at reviewed conferences and published in refereed journals. In addition, we have contributed chapters to the volumes *Machine Vision: Algorithms, Architectures and Systems* [71], *Advances in Machine Vision* [72], *Machine Vision for Three-Dimensional Scenes* [73] and *Analysis and Interpretation of Range Images* [74]. We were invited to present our research on multisensor fusion at the National Science Foundation Workshop on Range Image Processing [8] and the Workshop on Machine Vision for Three Dimensional Scenes [73] as well as several NATO Advanced Research Workshops [6], [7], [67].

One particular measure of the impact of our research in synergetic multisensor fusion is the outstanding success of the NATO Advanced Research Workshop on Multisensor Fusion for Computer Vision, held in Grenoble, France in June 1989 (Director: J. K. Aggarwal). Recognized leaders in the academic, governmental, or industrial research communities around the world met to discuss the latest advances in the principles and issues in multisensor fusion, information fusion for navigation, multisensor fusion for object recognition, network approaches to multisensor fusion, computer architectures for multisensor fusion, and applications of multisensor fusion.

In addition to the publications cited in the research detailed above, several other papers are forthcoming. "Sensor Data, Analysis, and Fusion," will appear in the book *Encyclopedia of Artificial Intelligence*, edited by Prof. A. Rosenfeld, to be published by John Wiley and Sons. "Sensor Data Fusion for Robotic Systems," will appear in the volume *Advances in Robotic System Dynamics and Control Systems*, edited by Dr. C. T. Leondes of the University of Washington, to be published by Academic Press.

We have made much progress in synergetic multisensor fusion, but much work remains to be done towards the development of truly general-purpose computer vision systems to reach the ultimate goal of computer vision research, which is to develop and engineer machines with the ability to sense, understand, and act upon their environments in an autonomous manner. Toward that end, it is the recommendation of the principal investigator and other members of our research team that the research in multisensor computer vision be continued.

Bibliographic References.

[1] N. Nandhakumar and J.K. Aggarwal, "Integrating Information from Thermal and Visual Images for Scene Analysis", in *Proc. of SPIE*, Vol. 635, 1986, pp 132 - 142.

[2] N. Nandhakumar and J.K. Aggarwal, "Thermal and Visual Information Fusion for Outdoor Scene Perception", Proc. *IEEE International Conference on Robotics and Automation*, Philadelphia, PA, April 24-29, 1988, pp. 1306-1308. [ARO 30]*

[3] N. Nandhakumar and J.K. Aggarwal, "Multisensor Fusion for Scene Perception - Integrating Thermal and Visual Imagery", Computer and Vision Research Center, The University of Texas at Austin, Tech. Report TR-87-9-41, August 1987.

[4] N. Nandhakumar and J. K. Aggarwal, "Integrated Analysis of Thermal and Visual Images for Scene Interpretation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 10, No. 4, pp. 469-481, July 1988. [ARO 51]

[5] N. Nandhakumar and J.K. Aggarwal, "Multisensor Integration - Experiments in Integrating Thermal and Visual Sensors", full paper in *Proc. First International Conference on Computer Vision*, London, England, June 8-11, 1987, pp. 83 - 92.

[6] N. Nandhakumar and J.K. Aggarwal, "A Phenomenological Approach to Thermal and Visual Sensor Fusion", presented at *NATO Advanced Research Workshop on Highly Redundant Sensing for Robotic Systems*, Il Ciocco, Italy, May 16-20, 1988. [ARO 28]

[7] J. K. Aggarwal, "Multisensor Computer Vision -- Issues and Results," *Proc. of the NATO Advanced Research Workshop on Multisensor Fusion for Computer Vision*, Grenoble, France, June 26-30, 1989. [ARO 4]

[8] N. Nandhakumar and J.K. Aggarwal, "Multisensor Fusion for Automatic Scene Interpretation - Research Issues and Directions", *Proc. NSF Workshop on Range Image Processing*, Ramesh Jain, Ed., March 1988. [ARO 29]

[9] C. Oh, N. Nandhakumar, and J. K. Aggarwal, "Integrated Modeling of Thermal and Visual Image Generation," *Proc.IEEE Conf.* on *Computer Vision and Pattern Recognition*, May 1989, pp. 356-362. [ARO 31]

[10] S. Karthik and J. K. Aggarwal, "Simultaneous Modeling of 3-D Objects for Visual and Thermal Image Synthesis," *Proc. SPIE Optical Engineering Southcentral '90 Regional Technical Symposium*, Dallas, Texas, May 1990. [ARO 45]

---

* [ARO ##] designates publications supported by the Army Research Office under this contract. Refer to list of publications and technical reports.

[11] Chen-Chau Chu, N. Nandhakumar, and J. K. Aggarwal, "Scene Segmentation and Information Integration from Laser Radar Range and Intensity Data," *Proc. 1988 Conference on Pattern Recognition for Advanced Missile Systems*, pp. 93-105. [ARO 18]

[12] Chen-Chau Chu, N. Nandhakumar, and J. K. Aggarwal, "Interpreting Segmented Laser Radar Images Using a Knowledge-Based System," *Proc. Sensor Fusion Workshop II: Human and Machine Strategies*, November 1989, Philadelphia, PA, pp. 314-323. [ARO 20]

[13] Chen-Chau Chu, N. Nandhakumar, and J. K. Aggarwal, "Image Segmentation Using Laser Radar Data, *Pattern Recognition*, Vol. 23, No. 6, 1990, pp. 569-581. [ARO 19]

[14] Chen-Chau Chu and J. K. Aggarwal, "Multi-Sensor Image Interpretation Using Laser Radar and Thermal Images," *Proc. 7th IEEE Conference on Artificial Intelligence Applications* (Miami Beach, February 1991), in press. [ARO 48]

[15] Chen-Chau Chu and J. K. Aggarwal, "The Interpretation of Laser Radar Images by a Knowledge-Based System," *Machine Vision and Applications*, in press. [ARO 49]

[16] Asar, H., N. Nandhakumar, and J. K. Aggarwal, "Pyramid-Based Image Segmentation Using Multisensory Data," *Pattern Recognition*, Vol. 23, No. 6, 1990, pp. 583-593. [ARO 10]

[17] C. C. Chu and J. K. Aggarwal, "The Integration of Region and Edge-Based Segmentation," *Proceedings of the IEEE International Conference on Computer Vision*, Osaka, Japan, December 1990, pp. 117-120. [ARO 42]

[18] B. C. Vemuri and J. K. Aggarwal, "3-Dimensional Reconstruction of Objects from Range Data," *Proc. of the 7th International Conference on Pattern Recognition (Montreal, July 1984)*, pp. 752-754.

[19] B.C. Vemuri, A. Mitiche and J.K. Aggarwal, "Curvature based representation of objects from range data", *Image and Vision Computing*, 4(2), May 1986, pp. 107-114.

[20] B. C. Vemuri, A. Mitiche, and J. K. Aggarwal, "3-D Object Representation from Range Data Using Intrinsic Surface Properties," Edited by T. Kanade, *Three-Dimensional Machine Vision*, Kluwer Academic Publishers, 1986, pp. 241-266.

[21] B. C. Vemuri and J. K. Aggarwal, "3-D Model Construction from Multiple Views Using Range and Intensity Data," *Proc. Computer Vision and Pattern Recognition Conf.*, Miami Beach, Florida, June 1986, pp. 435-437.

[22] B.C. Vemuri and J.K. Aggarwal, "Representation and recognition of objects from depth maps", *IEEE Trans. Circuits and Systems*, Vol. 34, No. 11, November 1987, pp. 1351-1363.

[23] J. P Brady, N. Nandhakumar, and J. K. Aggarwal, "Recent Progress in Object Recognition from Range Data," *Image and Vision Computing*, Vol. 7, No. 4, pp. 295-307, November 1989. [ARO 13]

[24] Brady, J.P., N. Nandhakumar and J.K. Aggarwal, "Recent Progress in the Recognition of Objects from Range Data", *Proc. Ninth International Conference on Pattern Recognition*, Rome, Italy, November 14-17, 1988, pp. 85-92. [ARO 14]

[25] Farshid Arman, Bikash Sabata, and J. K. Aggarwal, "Hierarchial Segmentation of 3-D Range Images," *Proceedings of the 1989 IEEE International Conference on Systems, Man, and Cybernetics*, Cambridge, Massachusetts, pp. 156-161, November 1989. [ARO 9]

[26] Sabata, B., F. Arman, and J. K. Aggarwal, "Segmentation of 3-D Range Images Using Pyramidal Data Structures," *Proceedings of the IEEE International Conference on Computer Vision*, Osaka, Japan, December 1990, pp.662-666. [ARO 41]

[27] F. Arman and J. K. Aggarwal, "Object Recognition in Dense Range Images Using a CAD System as a Model Base," *Proc. IEEE International Conference on Robotics and Automation*, Cincinnati, Ohio, May 1990, pp. 1858-1863. [ARO 8]

[28] C.H. Chien, Y.B. Sim and J.K. Aggarwal, "Generation of Volume Surface Octree from Range Data", *Proc. Computer Vision and Pattern Recognition Conf.*, Ann Arbor, MI, June 1988, pp.254-260. [ARO 17]

[29] J. K. Aggarwal and B. Sabata, "Motion from a Sequence of Range Images," *IEEE International Workshop on Intelligent Motion Control*, Istanbul, Turkey, August 1990. [ARO 2]

[30] C. H. Chien, and J. K. Aggarwal, "Model Construction and Shape Recognition from Occluding Contours, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 4, April 1989. [ARO 16]

[31] P. Besl and R. Jain, "Three-Dimensional Object Recognition," *ACM Computing Surveys*, Vol. 17, No. 1, March 1985, pp. 75-145.

[32] T. J. Fan, *Describing and Recognizing 3-D Objects Using Surface Properties*, Springer-Verlag, New York, New York, 1990.

[33] M. Brady, J. Ponce, A. Yullie, and H. Asada, "Describing Surfaces," *Computer Vision, Graphics, and Image Processing* Vol. 32, 1985, pp 1-28.

[34] R. Hoffman and A.K. Jain, "Segmentation and Classification of Range Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-9, No. 5, September 1987.

[35] P. Boulanger and M. Rioux, "Segmentation of Planar and Quadric Surfaces," *Intelligent Robots and Computer Vision: Sixth in Series*, D.P. Casasent and E.L. Hall, Eds., Proc. SPIE 848, Cambridge, Massachusetts, November 2-6, 1987, pp. 395-403

[36] J. Han, R. A. Voltz, and T. N. Mudge. "Range Image Segmentation and Surface Parameter Extraction for 3-D Object Recognition of Industrial Parts," *Proceedings of 1987 IEEE Conference on Robotics and Automation*, 1987, pp 380-386.

[37] P. J. Flynn and A. K. Jain, "Surface Classification: Hypothesis Testing and Planar Estimation," *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 1988, pp 261-267.

[38] Z. C. Lin, T. S. Huang, S. D. Blostein, H. Lee, and E. A. Margerum, "Motion Estimation from 3-D Point Sets with and without Correspondences," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Miami Beach, Florida, 1986.

[39] J. Aloimonos and I. Rigoutsos, "Determining the 3-D Motion of a Rigid Planar Patch without Correspondence," *Proc. IEEE Computer Society Workshop on Motion: Representation and Analysis*," May 1986, pp. 167-174.

[40] B. Sabata, N. Nandhakumar and J.K. Aggarwal, "Motion Estimation Using Surface Correspondences", *Proc. Optical Soc. of America Image Understanding and Machine Vision Topical Meeting*, Cape Cod, MA, June 12-14, 1989, pp. 50 - 53. [ARO 36]

[41] J.K. Aggarwal and N. Nandhakumar, "On the Computation of Motion from a Sequence of Images", invited paper in *Proceedings of the IEEE* , Vol. 76, No. 8, August 1988, pp. 917-935. [ARO 6]

[42] C. L. Cheng and J. K. Aggarwal, "A Two-Stage Hybrid Approach to the Correspondence Problem via Forward Searching and Backward-Correcting," *Proc. 10th International Conference on Pattern Recognition*, Atlantic City, New Jersey, June 1990, pp. 173-179. [ARO 15]

[43] Y. L. Chang and J. K. Aggarwal, "Reconstructing 3D Lines from a Sequence of 2D Projections: Representation and Estimation," *Proc. IEEE International Conference on Computer Vision*, Osaka, Japan, December 1990, pp.101-105. [ARO 43]

[44] Jeffrey J. Rodriguez and J. K. Aggarwal, "Navigation Using Image Sequence Analysis and 3-D Terrain Matching," *Proc. IEEE Workshop on Interpretation of 3D Scenes (November 1989, Austin Texas)*, pp. 200-207, 1990. [ARO 32]

[45] J.J. Rodriguez and J.K. Aggarwal, "Terrain Matching Using Image Sequence Analysis", Proc. *Optical Soc. of America Image Understanding and Machine Vision Topical Meeting*, Cape Cod, MA, June 12-14, 1989, pp. 30 - 33. [ARO 35]

[46] Rodriguez, J. J., and J. K. Aggarwal, "Terrain Matching by Analysis of Aerial Images," *Proceedings of the IEEE Internaticnal Conference on Computer Vision*, Osaka, Japan, December 1990, pp. 677-681. [ARO 39]

[47] J.J. Rodriguez and J. K. Aggarwal, "Matching Aerial Images to 3D Terrain Maps" *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 12, No. 12, 1990, pp. 1138-1149. [ARO 38]

[48] W. Eric L. Grimson, "Computational Experiments with a Feature Based Stereo Algorithm," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 7, pp. 17-34, January 1985.

[49] William Eric Grimson and Eric Leifur, *From Images to Surfaces.* Cambridge, MA: The MIT Press, 1981.

[50] D. Marr and E. Hildreth, "Theory of Edge Detection," *Proc. of the Royal Society of London*, Vol. B207, pp. 187-217, February 1980.

[51] W. N. Martin and J. K. Aggarwal, "Computer Analysis of Dynamic Scenes Containing Curvilinear Figures," *Pattern Recognition*, Vol 11, pp. 169-178, 1979.

[52] James W. McKee and J. K. Aggarwal, "Computer Recognition of Partial Views of Curved Objects, *IEEE Trans. on Computers*, Vol. 26, pp. 790-800, August 1977.

[53] Jeffrey J. Rodriguez and J. K. Aggarwal, "Quantization Error in Stereo Imaging," *Proc. IEEE-CS Conf. on Computer Vision and Pattern Recognition*, pp. 153-158, 1988. [ARO 34]

[54] Rodriguez, J. J. and J. K. Aggarwal, "Stochastic Analysis of Stereo Quantization Error," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 12, No. 5, May 1990, pp. 467-470. [ARO 33]

[55] U. Dhond and J.K. Aggarwal, "Closer Look at the Contribution of a Third Camera Toward Accuracy in Stereo Correspondence", *Proc. Optical Soc. of America Image Understanding and Machine Vision Topical Meeting*, Cape Cod, MA, June 12-14, 1989, pp. 78 - 81. [ARO 21]

[56] U. R. Dhond and J. K. Aggarwal, "Binocular vs. Trinocular Stereo," *Proc. IEEE International Conf. on Robotics and Automation*, Cincinnati, Ohio, May 1990, pp. 2045-2050. [ARO 23]

[57] U. R. Dhond and J. K. Aggarwal, "Cost-Benefit Analysis of a Third Camera in Stereo Correspondence," *International Journal of Computer Vision,* in press. [ARO 40]

[58] M. Ali, W.N. Martin and J.K. Aggarwal, "Color-based Computer Analysis of Aerial Photographs", *Computer Graphics and Image Processing,* Vol. 9, No. 3, March 1979, pp. 282-293.

[59] S.A. Underwood and J.K. Aggarwal, "Interactive Computer Analysis of Aerial Color Infrared Photographs", *Computer Graphics and Image Processing,* Vol. 6, No. 1, February 1977, pp. 1-24.

[60] M. Ali and J.K. Aggarwal, "Automatic Interpretation of Infrared Aerial Color Photographs of Citrus Tree Orchards having Infestations of Insect Pests and Diseases", *IEEE Trans. Geoscience Electronics,* Vol. 15, No. 3, July 1977, pp. 170-179.

[61] D.H. Williams and J.K. Aggarwal, "Computer Detection and Classification of Three Citrus Infestations", *Computer Graphics and Image Processing* (14), 1980, pp. 373-390.

[62] A. Sarabi and J.K. Aggarwal, "Segmentation of Chromatic Images", *Pattern Recognition,* Vol. 13, No. 6, 1981, pp. 417-427.

[63] D.C. Baker, J.K. Aggarwal and S.S. Hwang, "Geometry-guided Segmentation of Outdoor Scenes", *Proc. of SPIE Conf. Applications of Artificial Intelligence VI,* Vol. 937, Orlando, Florida, April 4-6, 1988, pp. 576-583. [ARO 12]

[64] D. C. Baker, S. S. Hwang, and J. K. Aggarwal, "Detection and Segmentation of Man-made Objects in Outdoor Scenes: Concrete-Bridges," *J. Opt. Soc. America A,* Vol. 6, No. 6, June 1989, pp. 938-950. [ARO 11]

[65] D. C. Baker, J. K. Aggarwal, and S. S. Hwang, "Geometry Guided Incremental Segmentation," *Proc. 1987 IEEE Computer Vision Workshop* (Miami Beach, Florida), N. Ahuja and T. Huang, Eds., pp. 237-239. [ARO 44]

[66] X. Lebegue, D. C. Baker, and J. K. Aggarwal, "2-D and 3-D Model-Based Recognition of Man-made Objects in Outdoor Scenes," *Proc. 6th Scandanavian Conf. on Image Analysis,* Oulu, Finland, 1989. [ARO 24]

[67] X. Lebegue and J. K. Aggarwal, "Fusion of Color and Geometric Information," *Proc. of the NATO Advanced Research Workshop on Multisensor Fusion for Computer Vision,* Grenoble, France, June 26-30, 1989, J. K. Aggarwal, Editor, Springer-Verlag, in press. [ARO 25]

[68] R. Talluri and J.K.Aggarwal, "Positional estimation for a mobile robot in an outdoor environment," *IEEE Workshop on Intelligent Robots and Systems,* IROS '90, pp. 159-166, Japan, July 90. [ARO 37]

[69] R. Talluri and J.K.Aggarwal, "A Positional estimation technique for and autonomous land vehicle in an unstructured environment," i-SAIRAS '90, Japan, Nov 90. [ARO 52]

[70] R. Talluri and J.K.Aggarwal,"Edge Visibility Regions - a new representation of the environment of a mobile robot," IAPR Workshop on Machine Vision Applications, MVA '91) pp. 375-380,Japan, Nov 90. [ARO 53]

[71] Y. F. Wang and J K. Aggarwal, "Inference of Object Surface Structure From Structured Lighting - An Overview," in *Machine Vision, Algorithm, Architectures and Systems,* (Vol. 20 in the series, *Perspectives in Computing*) Edited by H. Freeman, Academic Press, Inc., 1988, pp. 193-220. [ARO 54]

[72] C. H. Chien and J K. Aggarwal, "3-D Structure from 2-D Images," *Advances in Machine Vision,* Edited by Jorge L. C. Sanz, (Springer Series in Perception Engineering), Springer-Verlag, 1989, pp. 64-121. [ARO 55]

[73] J. K. Aggarwal, "Segmentation and Analysis of Multi-Sensor Images," in *Machine Vision for Three Dimensional Scenes,* edited by H. Freeman, Academic Press, Inc., 1990, pp. 267-299. [ARO 56]

[74] J. K. Aggarwal and N. Nandhakumar, "Multisensor Fusion for Automatic Scene Interpretation," in *Analysis and Interpretation of Range Images,* R. C. Jain and A. K. Jain, Eds., Springer-Verlag Press, pp. 339-361, 1990. [ARO 5]

SYNERGETIC MULTISENSOR FUSION
Publications and Technical Reports

1. Aggarwal, J. K. "Multisensor Fusion for Computer Vision," presented at the NATO Advanced Research Workshop on Sensory Systems for Robotic Control Closing Workshop, Il Ciocco, Tuscany, Italy, November 1989.
2. Aggarwal, J. K. and B. Sabata, "Motion from a Sequence of Range Images," *IEEE International Workshop on Intelligent Motion Control*, Istanbul, Turkey, August 1990.
3. Aggarwal, J. K., "Issues in AI in Computer Vision," introductory presentation at the Work-shop on Artificial Intelligence in Computer Vision, held in conjunction with the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, California, June 5, 1989.
4. Aggarwal, J. K., "Multisensor Computer Vision -- Issues and Results," *Proc. of the NATO Advanced Research Workshop on Multisensor Fusion for Computer Vision*, Grenoble, France, June 26-30, 1989.
5. Aggarwal, J. K.and N. Nandhakumar, "Multisensor Fusion for Automatic Scene Interpretation," in *Analysis and Interpretation of Range Images*, R. C. Jain and A. K. Jain, Eds., Springer-Verlag Press, pp. 339-361, 1990.
6. Aggarwal, J.K. and N. Nandhakumar, "On the Computation of Motion from a Sequence of Images", invited paper in *Proceedings of the IEEE* , Vol. 76, No. 8, August 1988, pp. 917-935.
7. Aggarwal, J.K. "Structure and Motion from Images", Office of Naval Research sponsored *Indo-US Workshop on Systems and Signal Processing*, Bangalore, India, Jan 8-12, 1988, pp 9-10.
8. Arman, F. and J. K. Aggarwal, "Object Recognition in Dense Range Images Using a CAD System as a Model Base," *Proceedings of the IEEE International Conference on Robotics and Automation*, Cincinnati, Ohio, May 1990, pp. 1858-1863.
9. Arman, Farshid, Bikash Sabata, and J. K. Aggarwal, "Hierarchial Segmentation of 3-D Range Images," *Proceedings of the 1989 IEEE International Conference on Systems, Man, and Cybernetics*, Cambridge, Massachusetts, pp. 156-161, November 1989.
10. Asar, H., N. Nandhakumar, and J. K. Aggarwal, "Pyramid-Based Image Segmentation Using Multisensory Data," *Pattern Recognition*, Vol. 23, No. 6, 1990, pp. 583-593.
11. Baker, D. C., S. S. Hwang, and J. K. Aggarwal, "Detection and Segmentation of Man-made Objects in Outdoor Scenes: Concrete Bridges," *Journal of the Optical Society of America*, vol. 6, no. 6, June 1989, pp. 938-950.
12. Baker, D.C., J.K. Aggarwal and S.S. Hwang, "Geometry Guided Segmentation of Outdoor Scenes", *Proc. of SPIE Conf. on Applications of Artificial Intelligence VI,* Orlando, Florida, April 4-6, 1988, pp. 576-583.
13. Brady, J. P., N. Nandhakumar, and J. K. Aggarwal, "Recent Progress in Object Recognition from Range Data," *Image and Vision Computing*, Vol. 7, No. 4, pp. 295-307, 1989.
14. Brady, J.P., N. Nandhakumar and J.K. Aggarwal, "Recent Progress in the Recognition of Objects from Range Data", *Proc. Ninth International Conference on Pattern Recognition*, Rome, Italy, November 14-17, 1988, pp. 85-92.

15. Cheng, C. L. and J. K. Aggarwal, "A Two-Stage Hybrid Approach to the Correspondence Problem via Forward Searching and Backward-Correcting," *Proceedings of the 10th International Conference on Pattern Recognition*, Atlantic City, New Jersey, June 1990, pp. 173-179.

16. Chien, C. H. and J. K. Aggarwal, "Model Construction and Shape Recognition from Occluding Contours, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 4, April 1989.

17. Chien, C.H., Y.B. Sim, and J.K. Aggarwal, "Generation of Volume/Surface Octree from Range Data", *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Ann Arbor, MI, June 5-9, 1988, pp. 254-260.

18. Chu, Chen-Chau, N. Nandhakumar, and J. K. Aggarwal, "Scene Segmentation and Information Integration from Laser Radar Range and Intensity Data," *Proceedings of the 1988 Conference on Pattern Recognition for Advanced Missile Systems*, Redstone Arsenal, Huntsville, Alabama, pp. 93-105. (Conf. November 1988) Published June 1989.

19. Chu, Chen-Chau, N. Nandhakumar, and J. K. Aggarwal, "Image Segmentation Using Laser Radar Data, *Pattern Recognition*, Vol. 23, No. 6, 1990, pp. 569-581.

20. Chu, Chen-Chau, N. Nandhakumar, and J. K. Aggarwal, "Interpreting Segmented Laser Radar Images Using a Knowledge-Based System," *Proceedings of the Sensor Fusion Workshop II: Human and Machine Strategies*, November 1989, Philadelphia, PA, pp. 314-323.

21. Dhond, U. and J.K. Aggarwal, "Closer Look at the Contribution of a Third Camera Toward Accuracy in Stereo Correspondence", Proc. *Optical Soc. of America Image Understanding and Machine Vision Topical Meeting*, Cape Cod, MA, June 12-14, 1989, pp. 78 - 81.

22. Dhond, U. R. and J. K. Aggarwal, "Structure from Stereo -- A Review," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, No. 6, pp. 1489-1510, November/December 1989.

23. Dhond, U. R. and J. K. Aggarwal, "Binocular vs. Trinocular Stereo," *Proceedings of the IEEE International Conference on Robotics and Automation*, Cincinnati, Ohio, May 1990, pp. 2045-2050.

24. Lebegue, X., D. C. Baker, and J. K. Aggarwal, "2-D and 3-D Model-Based Recognition of Man-made Objects in Outdoor Scenes," *Proc. 6th Scandanavian Conference on Image Analysis*, Oulu, Finland, June 19-22, 1989.

25. Lebegue, X., "Fusion of Color and Geometric Information," *Proc. of the NATO Advanced Research Workshop on Multisensor Fusion for Computer Vision*, Grenoble, France, June 26-30, 1989.

26. Mitiche, A., O. Faugeras, and J. K. Aggarwal, "Counting Straight Lines," *Computer Vision, Graphics, and Image Processing* 47, 353-360, 1989.

27. Nandhakumar, N. " Application of AI in Multisensory Computer Vision", invited presentation at *IEEE Computer Society Workshop on Artificial Intelligence for Computer Vision*, San Diego, CA, June 5, 1989.

28. Nandhakumar, N. and J.K. Aggarwal, "A Phenomenological Approach to Thermal and Visual Sensor Fusion", presented at *NATO Advanced Research Workshop on Highly Redundant Sensing for Robotic Systems*, Il Ciocco, Italy, May 16-20, 1988.

29. Nandhakumar, N. and J.K. Aggarwal, "Multisensor Fusion for Automatic Scene Interpretation - Research Issues and Directions", to appear in Proc. *NSF Workshop on Range Image Processing*, (Ed.) Ramesh Jain, March 1988.

30. Nandhakumar, N. and J.K. Aggarwal, "Thermal and Visual Information Fusion for Outdoor Scene Perception", *Proc. of IEEE International Conference on Robotics and Automation*, Philadelphia, PA, April 25-29, 1988, pp. 1306-1308.

31. Oh, C.H., N. Nandhakumar and J.K. Aggarwal, "Integrated Modelling of Thermal and Visual Image Generation", *Proc. IEEE Computer Vision and Pattern Recognition Conf.*, San Diego, CA, June 4-6, 1989, pp. 356 - 362.

32. Rodriguez, J. J. and J. K. Aggarwal, "Navigation Using Image Sequence Analysis and 3-D Terrain Matching," *Proceedings of the IEEE Workshop on the Interpretation of 3-D Scenes*, Austin, Texas, pp. 200-207, November 1989.

33. Rodriguez, J. J. and J. K. Aggarwal, "Stochastic Analysis of Stereo Quantization Error," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 5, May 1990, pp. 467-470.

34. Rodriguez, J.J. and J.K. Aggarwal, "Quantization Error in Stereo Imaging", *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Ann Arbor, MI, June 5-9, 1988, pp. 153-158.

35. Rodriguez, J.J. and J.K. Aggarwal, "Terrain Matching Using Image Sequence Analysis", Proc. *Optical Soc. of America Image Understanding and Machine Vision Topical Meeting*, Cape Cod, MA, June 12-14, 1989, pp. 30 - 33.

36. Sabata, B. , N. Nandhakumar and J.K. Aggarwal, "Motion Estimation Using Surface Correspondences", Proc. *Optical Soc. of America Image Understanding and Machine Vision Topical Meeting*, Cape Cod, MA, June 12-14, 1989, pp. 50 - 53.

37. Talluri, R. and J. K. Aggarwal, "Position Estimation for a Mobile Robot in an Unstructured Environment," *Proceedings of the IEEE International Workshop on Intelligent Robots and Systems,"* Tsuchiura, Japan, July 1990.

38. Rodriguez, J. J., and J. K. Aggarwal, "Matching Aerial Images to 3D Terrain Maps" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 12, 1990, pp. 1138-1149.

39. Rodriguez, J. J., and J. K. Aggarwal, "Terrain Matching by Analysis of Aerial Images," *Proceedings of the IEEE International Conference on Computer Vision*, Osaka, Japan, December 1990, pp.677-681.

40. Dhond, U. R. and J. K. Aggarwal, "Cost-Benefit Analysis of a Third Camera in Stereo Correspondence," *International Journal of Computer Vision*, in press.

41. Sabata, B., F. Arman, and J. K. Aggarwal, "Segmentation of 3-D Range Images Using Pyramidal Data Structures," *Proceedings of the IEEE International Conference on Computer Vision*, Osaka, Japan, December 1990, pp.662-666.

42. Chu, C. C., and J. K. Aggarwal, "The Integration of Region and Edge-Based Segmentation," *Proceedings of the IEEE International Conference on Computer Vision*, Osaka, Japan, December 1990, pp. 117-120.

43. Chang, Y. L., and J. K. Aggarwal, "Reconstructing 3D Lines from a Sequence of 2D Projections: Representation and Estimation," *Proceedings of the IEEE International Conference on Computer Vision*, Osaka, Japan, December 1990, pp.101-105.

44. Baker, D. C., J. K. Aggarwal, and S. S. Hwang, "Geometry Guided Incremental Segmentation," *Proc. IEEE Computer Vision Workshop* (Miami Beach, Florida, December 1987), N. Ahuja and T. Huang, Editors, pp. 237-239.

45. Karthik, S. and J. K. Aggarwal, "Simultaneous Modeling of 3-D Objects for Visual and Thermal Image Synthesis," *Proc. of the Optical Engineering Southcentral '90 Regional Technical Symposium,* Dallas, Texas, May 1990.

46. Brady, J.P., N. Nandhakumar and J.K.Aggarwal, "Recent progress in the recognition of objects from range data", Tech. Report TR-88-1-46, Computer and Vision Research Center, The University of Texas at Austin, Austin, Texas, January 1988

48. Chu, Chen-Chau and J. K. Aggarwal, "Multi-Sensor Image Interpretation Using Laser Radar and Thermal Images," *Proc. 7th IEEE Conference on Artificial Intelligence Applications* (Miami Beach, February 1991), in press.

49. Chu, Chen-Chau and J. K. Aggarwal, "The Interpretation of Laser Radar Images by a Knowledge-Based System," *Machine Vision and Applications,* in press.

50. Wang, Y. F., and J. K. Aggarwal, "Integration of Active and Passive Sensing Techniques for Representing Three-Dimensional Objects," *IEEE Trans. Robotics and Automation,* Vol. 5, No. 4., 1989, pp. 460-471.

51. Nandhakumar, N. and J. K. Aggarwal, "Integrated Analysis of Thermal and Visual Images for Scene Interpretation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 10, No. 4, July 1988, pp. 469-481.

52. R. Talluri and J.K.Aggarwal, "A Positional estimation technique for an autonomous land vehicle in an unstructured environment," i-SAIRAS '90, Japan, Nov 90.

53. R. Talluri and J.K.Aggarwal,"Edge Visibility Regions - a new representation of the environment of a mobile robot," *Proc. IAPR Workshop on Machine Vision Applications,* (Japan, Nov 90), pp. 375-380.

54. Y. F. Wang and J K. Aggarwal, "Inference of Object Surface Structure From Structured Lighting - An Overview," in *Machine Vision, Algorithm, Architectures and Systems,* (Vol. 20 in the series, *Perspectives in Computing*) Edited by H. Freeman, Academic Press, Inc., 1988, pp. 193-220.

55. C. H. Chien and J K. Aggarwal, "3-D Structure from 2-D Images," *Advances in Machine Vision,* Edited by Jorge L. C. Sanz, (Springer Series in Perception Engineering), Springer-Verlag, 1989, pp. 64-121 .

56. Aggarwal, J. K., "Segmentation and Analysis of Multi-Sensor Images," in *Machine Vision for Three Dimensional Scenes,* edited by H. Freeman, Academic Press, Inc., 1990, pp. 267-299.

Scientific Personnel Supported By and Degrees Awarded During This Project

Prof. J.K. Aggarwal,	Principal Investigator	
Dr. Hong-Qian Lu	Research Associate	
Mr. F. May	Research Associate	
Dr. N. Nandhakumar	Research Associate	
Mr. F. Arman	Ph.D. Student	
Mr. H. Asar	M.S. Electrical Engineering	August 1988
Mr. D.C. Baker	Ph.D Electrical Engineering	May 1989
Mr. Y. L. Chang	M.S. Student	
Mr. C. L. Cheng	M.S. Electrical Engineering	December 1989
Mr. R.J. Chou	Ph.D. Student	
Mr. C.C. Chu	Ph.D. Student	
Mr. U. Dhond	Ph.D. Student	
Mr. S. Karthik	M.S. Student	December 1990
Ms. K. Kumari	M.S. Electrical Engineering	August 1990
Mr. Jason Ng	B.S. Student	May 1990
Mr. C.H. Oh	M.S. Student	May 1989
Mr. J. Rodriguez	Ph.D. Student	May 1990
Mr. B. Sabata	Ph.D. Student	
Mr. Matthew Schneble	M.S. Student	
Mr. Y. B. Sim	M.S. Student	May 1990
Mr. M. H. Sunwoo	Ph.D. Student	August 1990
Mr. Raj Talluri	Ph.D. Student	
Mr. Joseph Yip	Ph.D. Student	

Figures and Tables

Interpretation of Thermal and Visual Sensor Information

Segmentation and Understanding of Ladar Images

Interpretation of Range Imagery

• Segmentation of Range Images

• Motion from a Sequence of Range Images

Passive Aerial Navigation by Image Sequence Analysis